

VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA
EKONOMICKÁ FAKULTA

KATEDRA SYSTÉMOVÉHO INŽENÝRSTVÍ A INFORMATIKY

Aplikace umělé inteligence poskytující relevantní informaci podle profilu uživatele

An Artificial Intelligence Application Providing Relevant Information
According to User Profile

Student: Ing. Ondřej Zulák

Vedoucí diplomové práce: RNDr. Jaroslav Teda, Ph.D.

Ostrava 2012

Zadání diplomové práce

Student: **Ing. Ondřej Zulák**

Studijní program: N6209 Systémové inženýrství a informatika

Studijní obor: 6209T025 Systémové inženýrství a informatika

Specializace: 00 Systémové inženýrství a informatika

Téma: Aplikace umělé inteligence poskytující relevantní informaci podle
profilu uživatele
An Artificial Intelligence Application Providing Relevant Information
According to User Profile

Zásady pro vypracování:

1. Úvod
 2. Teorie neuronových sítí
 3. Analýza problematiky, specifikace procesů a datových toků
 4. Systémový návrh modelu
 5. Studie proveditelnosti
 6. Závěr
- Seznam použité literatury
Seznam zkratk
Prohlášení o využití výsledků diplomové práce
Seznam příloh
Přílohy

Seznam doporučené odborné literatury:

RUSSELL, Stuart. *Artificial Intelligence: A Modern Approach*. 3rd ed. Prentice Hall 2009, p. 1152. ISBN 978-0136042594.


VONDRÁK, Ivo. *Umělá inteligence a neuronové sítě*. 3. vyd. Ostrava: VŠB-TUO, 2009. 140 s. ISBN 978-80-248-1981-5.


Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **RNDr. Jaroslav Teda, Ph.D.**

Datum zadání: 25.11.2011

Datum odevzdání: 27.04.2012


Ing. Eva Moravcová, CSc.
vedoucí katedry


prof. Dr. Ing. Dana Dluhošová
děkanka fakulty



Místopřísežné prohlášení

Prohlašuji, že jsem celou práci včetně všech příloh vypracoval samostatně.

V Ostravě dne 27. 4. 2012

.....
Ing. Ondřej Zulák

Obsah

1. Úvod	1
2. Teorie neuronových sítí	3
2.1 Historie neuronových sítí	4
2.2 Charakteristiky neuronových sítí	6
2.3 Perceptronová síť	11
2.3.1 Organizační dynamika.....	12
2.3.2 Aktivní dynamika	13
2.3.3 Adaptivní dynamika	14
2.4 Kohonenova samoorganizující se mapa	16
2.4.1 Organizační dynamika.....	16
2.4.2 Aktivní dynamika	17
2.4.3 Adaptivní dynamika	18
2.5 Síť MADALINE.....	18
2.5.1 Organizační dynamika.....	18
2.5.2 Aktivní dynamika	18
2.5.3 Adaptivní dynamika	19
3. Analýza problematiky, specifikace procesů a datových toků.....	20
3.1 Analýza problematiky	20
3.2 Specifikace stávajícího informačního systému	22
4. Systémový návrh modelu	30
4.1 Výběr sítě	31
4.2 Model sítě.....	33
4.2.1 První podsíť	35
4.2.2 Druhá podsíť	37
4.3 Učení sítě.....	40
5. Studie proveditelnosti	44
5.1 Způsob prezentace neuronové sítě uživateli.....	47
5.2 Rizika projektu	49
6. Závěr	51
Seznam použité literatury	53
Seznam zkratk	
Prohlášení o využití výsledků diplomové práce	
Seznam příloh	

1. Úvod

Využití výpočetní techniky v naší společnosti i ve světě stále roste. Potenciál, kterým toto odvětví disponuje, je ohromný. Kromě hardware, který se neustále zmenšuje a zrychluje, nikoho již nepřekvapí, že počítačové programy umí vypočítat daně, umožnit komunikaci mezi lidmi, kteří jsou dokonce na opačné straně zeměkoule, apod. Naše nároky se zvyšují společně s novými možnostmi, které IT přináší. Většina dnešních počítačových programů je založena na procedurálních programovacích jazycích. Jinými slovy řečeno, programy, které užíváme, mají popsany přesný algoritmus, podle kterého mají postupovat. Tyto programy vytvářejí lidé, kteří znají danou problematiku a jsou schopni formulovat kroky vedoucí k řešení. Takto vytvořené programy jsou dobrými sluhý, udělají spousty práce za nás, ovšem možnosti jejich aplikací jsou omezené. Omezenost spočívá v tom, že nemohou řešit např. nestandardní úlohu, která algoritmem není zadána. Jsou schopny řešit jen maximálně to, co jim programátor určil. Klasické programy tuto hranici nikdy nepřekročí. Ovšem v našich životech existují situace, ve kterých nikdo přesně neví, jakým způsobem je řešit. Příkladem může být predikce počasí, zkoumání vazeb DNA, zjištění podobností mezi kriminálními případy.

Není důvod být z této situace sklíčený. Na složitějších programech dnes nepracuje jen jeden programátor. Vývoj software je v některých podnicích řízeným procesem, takže se na jeho vývoji podílí více lidí, počínaje systémovým analytikem a konče klientem. Všichni tito lidé přispívají svými znalostmi k řešení daného software. Nicméně i zde platí výše zmíněné omezení. Program opět odráží pouze znalosti lidí, které jsou zachyceny v algoritmu programu. Již dříve si lidé uvědomili, že je třeba vytvořit jiný přístup. Vzniklo tak odvětví umělé inteligence.

Umělá inteligence má své základy v 50. letech 20. století. Vývoj měl svá pozitivní i negativní období. Dnes umělá inteligence sice ještě netvoří běžnou součást našeho programového vybavení, ale je již s úspěchy využívána k řešení konkrétních problémů, jako například zmiňovaná předpověď počasí. Umělá inteligence v sobě zahrnuje více částí, kdy jednou z nich jsou neuronové sítě, kterými se zabývá tato diplomové práce. Odvětví je charakteristické tím, že se snaží prolomit omezení procedurálních jazyků. Expertní systém či neuronová síť tak jde za hranici poznatků lidí, kteří daný model umělé inteligence vytvořili. Tyto, můžeme říkat programy, jsou schopny obecně pracovat s neúplnými daty, velkým množstvím dat a pravidel,

umí najít souvislosti mezi vzory, učit se, měnit své nastavení v závislosti na čase a dokonce i zapomínat. Umělá inteligence vznikla spojením biologie a IT. Tak jako v lidském organismu může docházet k učení, dedukci apod., snaží se umělá inteligence napodobit tyto pochody uměle s použitím výpočetní techniky. Programy simulují myšlení a jejich chování se blíží chování lidskému.

Dalším prvkem, kterého se diplomová práce týká, jsou informace. Informace jsou důležitou součástí našich životů a zejména v podnikatelské sféře tvoří konkurenční výhodu. Nelze je zaměňovat s daty, které jsou na nižší úrovni a pouze popisují určitý jev. Diplomová práce se zabývá tím, jak zpracovat data, aby poskytla relevantní informaci. Ta má pro člověka větší hodnotu a je-li včasná a kvalitní, může například zvýšit podniku podíl na trhu, ověřit si marketingový průzkum, zjistit hrozby a příležitosti a podobně.

Cílem diplomové práce je prozkoumat možnosti neuronových sítí a navrhnout tak model neuronové sítě, který by dokázal uživateli doporučit relevantní informaci podle jeho profilu. Model bude popsán teoreticky s tím, že se posléze může prakticky realizovat v již existujícím informačním systému, ze kterého může čerpat data v podobě zpravodajských článků.

Po úvodu následuje kapitola, která popisuje nejdříve historii neuronových sítí. Pokračuje popis společných charakteristik neuronových sítí, jako je jednoduchý model neuronu, způsob učení neuronové sítě, vlastnosti tréninkové množiny, apod. Podkapitoly tvoří popis tří nepoužívanějších neuronových sítí, kterými jsou Perceptronová síť, Kohonenova samoorganizující se mapa a síť MADALINE.

Třetí kapitola popisuje problematiku, kterou se neuronová síť bude zabírat. Je zde uveden i informační systém, který bude neuronové síti poskytovat data.

Kapitola čtvrtá chronologicky popisuje činnosti vedoucí ke tvorbě výsledné neuronové sítě. Nejprve je stanovena konkrétní úloha. Následně je pro úlohu vybrán vhodný typ neuronové sítě. Dále kapitola obsahuje kompletní popis navrženého modelu neuronové sítě a celá jedna podkapitola je věnována procesu adaptace.

Pátá kapitola popisuje mimo jiné hardwarové a softwarové nároky neuronové sítě. Uvádí možný způsob prezentace neuronové sítě a poukazuje na rizika projektu.

2. Teorie neuronových sítí

„V současnosti patří neuronové sítě mezi významnou část počítačově orientované umělé inteligence, kde zaujaly postavení univerzálního matematicko-informatického přístupu ke studiu a modelování procesů učení. Kromě umělé inteligence mají neuronové sítě nezastupitelné uplatnění také v kognitivní vědě, lingvistice, neurovědě, řízení procesů, přírodních a společenských vědách, kde se pomocí nich modelují nejen procesy učení a adaptace, ale i široké spektrum různých problémů klasifikace objektů a také problémů řízení složitých průmyslových systémů. Neuronové sítě lze proto použít na řešení velkého množství úloh z oblasti klasifikace, predikce, optimalizace apod.“ (Volná, 2009, s. 658).

Marvin Minský uvádí, že: „Umělá inteligence je věda o vytváření strojů využívajících postupy, které prováděny člověkem, by vyžadovaly jeho inteligenci.“ Vondrák (2009, s. 7).

Neuronové sítě tvoří jednu samostatnou oblast umělé inteligence a mají své vlastnosti, které uvádí Vondrák (2009):

- „Neuronové sítě jsou inspirovány biologickými neuronovými sítěmi. Tato vlastnost určitým způsobem předurčuje, že uměle vytvořené neuronové sítě by měly být schopny, z hlediska základních principů, se chovat stejně nebo alespoň podobně jako jejich biologické vzory. Je zřejmé, že vytvoření umělého lidského mozku se všemi jeho schopnostmi je věc jen velmi těžce řešitelná ať už z hlediska kvantity jeho neuronů či jejich způsobů propojení, chování jednotlivých typů neuronů apod. Nicméně skýtá se tu šance simulovat alespoň některé funkce lidského myšlení a tyto pak implementovat.“ (Vondrák, 2009, s. 87).
- Paralelní zpracování informace je dalším rysem, který znamená, že zpracování neuronové sítě probíhá prostřednictvím celé sítě najednou.
- Znalosti, resp. jejich důležitost, je reprezentována vazbou mezi jednotlivými neurony. Pokud některá vazba vede ke správné odpovědi, pak je tato vazba posilována a naopak.
- Neexistuje algoritmus výpočtu, naopak vyskytuje se zde učení neuronových sítí. V současných programovacích jazycích je nutné zapsat algoritmus pro dosažení cíle. U neuronových sítí tomu tak není. Neuronovým sítím je předkládána tzv. trénovací množina, na základě které se síť učí a dá se říci, že si onen algoritmus vytváří sama.

Odpadá zde potřeba algoritmizace úlohy, takže pomocí neuronových sítí můžeme řešit i úlohy, u kterých algoritmus sami neznáme.

2.1 Historie neuronových sítí

Dle Šíma a Neruda (1996) se za zakladatele neuronových sítí považují Američané W. McCulloch a W. Pitts, kteří v roce 1943 vytvořili první matematický model neuronu. Jejich model byl pokusem převést biologický neuron na matematický model, který by realizoval aritmetické a logické funkce, což se jim i povedlo. Jejich objev neměl praktického využití, ovšem jak se ukázalo později, model vytvořil základy pro současné neuronové sítě.

Jedním z následníků byl v roce 1949 Donald Hebb, který vytvořil učící pravidlo pro mezineuronové rozhraní tzv. synapse. Inspiroval se podmíněnými reflexy u zvířat, kde předpokládal, že tyto reflexy jsou již vlastností jednotlivých neuronů. (Šíma a Neruda, 1996).

Marvin Minský o několik let později zkonstruoval dokonce první neuropočítač, který uměl automaticky adaptovat váhy synapsí, ovšem ačkoliv by se tak mohlo zdát, své uplatnění nenašel. (Šíma a Neruda, 1996).

V roce 1957 byl vynalezen tzv. perceptron F. Rosenblattem. Ve své podstatě se jednalo o zobecnění již zmíněného McCullochova a W. Pittsova modelu neuronu, ovšem s tím rozdílem, že perceptron již uměl pracovat s reálnými čísly. Jeho práce také zahrnovala výpočet učícího algoritmu, o kterém matematicky dokázal, že po určité době nalezne správný váhový vektor parametrů a to nezávisle na počátečním nastavení. Na základech perceptronu byl vytvořen další neuropočítač se jménem Mark I Perceptron. Jednalo se o účelově zkonstruovaný neuropočítač, který demonstroval dosažené výsledky a možnosti využití perceptronu. Neuropočítač rozpoznával znaky ze světelné mřížky velikosti 20 x 20 bodů. Perceptron vzbudil velká očekávání nejenom v řadách vědců. (Šíma a Neruda, 1996).

Téměř ve stejné době přichází B. Widrow se svými studenty s dalším typem neuronového výpočetního prvku nazvaným ADALINE. B. Widrow šel však ještě dále. Nejenže také demonstroval funkčnost na typových příkladech, ale také založil první komerční firmu, která

prodávala neuropočítače a jejich komponenty. Následovali další např. R. Barron, L. Gilstrap a jiní. V 60. letech 20. století se tak konečně dostavil rozmach neuronových sítí. Bohužel tento rozkvět s sebou přinesl i negativa, která v následujících letech převládla a odsunula neuronové sítě do pozadí. Z historického hlediska byla nalezena dvě největší negativa. Prvním negativem byl experimentální přístup k neuronovým sítím, který zanedbával analytický výzkum. Druhým byl pak neopodstatněně přehnaný optimismus, který vzbudil u veřejnosti vysoká očekávání a v závěru pak zklamal. V 60. letech 20. století se totiž prohlašovalo, že např. v krátkém čase vyvinou umělý mozek. Tento výrok je nereálný i v současné době. Pokud by ochabnul pouze zájem veřejnosti, nebylo by to tak zlé. Skutečnost byla však taková, že na výzkum a vývoj neuronových sítí byly dávány dotace. Krizi neuronových sítí tak využili jiní vědci, mimo jiné i Minský, který vytvořil první neuropočítač, aby převedli tyto dotace do jiných oblastí umělé inteligence. Záměr se jim nakonec podařil a tak v období mezi lety 1967 a 1982 neprobíhal v USA výzkum neuronových sítí skoro vůbec. (Šíma a Neruda, 1996).

V roce 1986 byl Davidem Rumelhartem, Geoffreyem Hintonem a Rolandem Williamsem publikován tzv. backpropagation algoritmus pro vícevrstvou neuronovou síť. Jak uvádí Vondrák (2009), jedná se o metodu zpětného šíření signálu v neuronové síti, která umožňuje její adaptaci. Byl to velký milník, protože se jednalo o průlomové řešení, které brzdilo pokrok. Bohužel jak se ukázalo později, metoda zpětného šíření signálu byla vynalezena již dříve, jen jí nebyl věnován dostatek pozornosti. Nicméně od 80. let 20. století se však situace změnila. Zejména v USA začaly do výzkumu opět proudit finance a jinde ve světě rovněž přicházeli na podstatné objevy. Z těch známějších je možno jmenovat např. Johna Hopfielda či Teuvo Kohonena. V roce 1987 již není pochyb o tom, že umělá inteligence, zejména pak neuronové sítě, budou mít budoucnost a tak vzniká první mezinárodní společnost pro výzkum neuronových sítí nazvaná International Neural Network Society (INNS). (Šíma a Neruda, 1996).

Rozvoj nejen neuronových sítí pokračuje dodnes. Společnost IBM dokonce v roce 2011 vyrobila první prototyp tzv. kognitivního čipu, který pracuje na podobném principu jako lidský mozek. I když používají technologii na bázi křemíku, nejsou předem naprogramovány, ale mohou se učit a reagovat na podněty z okolí. (Šmejkal, 2012, s. 58).

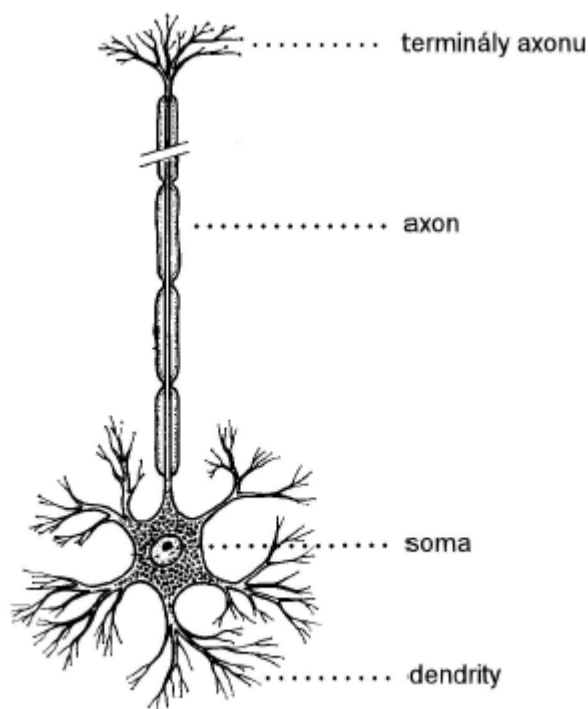
Budoucnost umělé inteligence není vůbec jasná. Daleko jsou doby, kdy se hovořilo se zarytou jistotou, že zanedlouho vytvoříme umělý lidský mozek, že budeme žít mezi roboty resp. kyborgy, kteří budou nerozpoznatelní od lidí apod. Je zřejmé, že lidský organismus je natolik propracovaný, že během několika let takové objevy nebudou možné. Otázkou však zůstává, zda je to vůbec reálné? Tímto se mimo jiné zabývá i Russell (2010), který odkazuje na již zjištěné poznatky především z oboru filozofie. Uvádí, že je nutné vyřešit problém duše a těla tzv. dualismus. Především se jedná o to, že kyborg si musí uvědomovat sám sebe. Je otázkou, jaké to uvědomění bude a lze předpokládat, že nebude ani podobné tomu lidskému. Vše totiž závisí na prostředí, ve kterém se my, resp. kyborg pohybuje a jak si uvědomuje své bytí v něm.

2.2 Charakteristiky neuronových sítí

Neuronové sítě vycházejí ze základního modelu neuronu, který byl inspirován biologickým vzorem. „Lidský mozek je složen asi z 10^{10} nervových buněk. Každý neuron může být propojen s jiným či dokonce s několika jinými. Vzniká tak obrovská síť vazeb. Biologický neuron má více krátkých výběžků (dendrity) a pouze jeden delší výběžek (axon). „Zatímco tělo neuronu je velké řádově mikrometry a dendrity několik milimetrů, axon může dosahovat délky desítek centimetrů až jednoho metru. Axon každého neuronu je zakončen tzv. synapsí (typicky jednou, ale může jich být i několik), která dosedá na jiný neuron. Přes synapse se přenášejí vzruchy mezi neurony; v důsledku chemických reakcí se v místě, kde synapse dosedá na neuron, mění propustnost buněčné membrány neuronu, tím se lokálně mění koncentrace kladných i záporných iontů vně a uvnitř buňky a tedy i membránový potenciál. Některé synaptické vazby mají charakter excitační (zvyšují membránový potenciál), jiné mají charakter inhibiční (snižují membránový potenciál). Dílčí účinky synaptických vazeb se na neuronu kumulují a ve chvíli, kdy celkový membránový potenciál přesáhne určitý práh, neuron je aktivován a přes svoji synapsi začne působit na další neurony, se kterými je spojen.“ (Berka, 2008, s. 116).

V této kapitole je uveden matematický model neuronu a jsou popsány základní termíny společné pro různé typy neuronových sítí. Jednotlivým typům sítí jsou věnovány další kapitoly.

Obr. 2.1 – Biologický neuron



Zdroj: Šíma a Neruda (1996)

Matematický model neuronu

Šíma a Neruda (1996, s. 24) definují model následovně. „Model má n obecně reálných vstupů x_1, \dots, x_n , které modelují dendrity. Vstupy jsou ohodnoceny odpovídajícími obecně reálnými synaptickými váhami w_1, \dots, w_n , které určují jejich propustnost. Ve shodě s neurofyzilogickou motivací mohou být synaptické váhy záporné, čímž se vyjadřuje jejich inhibiční charakter. Zvážená suma vstupních hodnot představuje vnitřní potenciál neuronu:“

$$\xi = \sum_{i=1}^n w_i \cdot x_i \quad (2.1)$$

Kde:

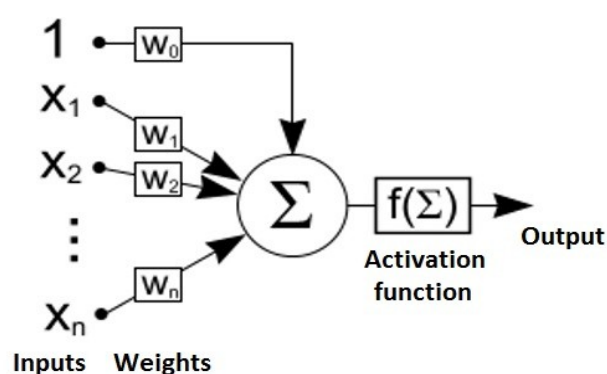
- n počet vstupů,
- w_i váha vstupu i ,
- x_i hodnota vstupu i .

Pokud hodnota **vnitřního potenciálu** značená ξ (na obr. 2.2 je znázorněn jako kruh se znakem suma) dosáhne tzv. prahové hodnoty, neuron se vybudí a indikuje výstup. Obdobně jako biologický neuron modeluje elektrický impuls axonu. O tom, zda je hodnota

vnitřního potenciálu překročena nebo ne, rozhoduje tzv. aktivační (přenosová) funkce, kterých existuje několik, např. ostrá nelinearita, lineární saturovaná funkce, standardní sigmoida či hyperbolický tangens. (Šíma a Neruda, 1996).

Na obr. 2.2 je možné vidět i jednotkový vstup s váhou w_0 . Tento vstup je označován jako **bias**, který „odpovídá váhové hodnotě přiřazené spojení mezi daným neuronem a fiktivním neuronem, jehož aktivace je vždy 1“. (Volná, 2008a, s. 40).

Obr. 2.2 – Matematický model neuronu (McCulloch, Pitts)



Zdroj: <http://encefalus.com/neurology-biology/neural-networks-real-neurons/>

Neuronová síť

„Neuronová síť se skládá z formálních neuronů, které jsou vzájemně propojené tak, že výstup neuronu je vstupem obecně více neuronů, podobně jako terminály axonu biologického neuronu jsou přes synaptické vazby spojeny s dendrity jiných neuronů. Počet neuronů a jejich vzájemné propojení v síti určuje tzv. architekturu (topologii) neuronové sítě. Z hlediska využití rozlišujeme v síti vstupní, pracovní (skryté, mezilehlé) a výstupní neurony. Stavby všech neuronů v síti určují tzv. **stav** neuronové sítě a synaptické váhy všech spojů představují tzv. **konfiguraci** neuronové sítě.“ (Šíma a Neruda, 1996, s. 29).

Neuronová síť se v čase vyvíjí, mění se stav neuronů i jednotlivé váhy. V souvislosti se změnou hovoříme o dynamice. Rozeznáváme tyto dynamiky (Šíma a Neruda, 1996):

- organizační,
- aktivní,
- adaptivní.

Organizační dynamika definuje topologie sítě, vymezuje její architekturu, případně její změnu. Aktivní dynamika řeší počáteční stav sítě a změny stavu sítě v čase při pevné konfiguraci a topologii. Dynamika je obvykle po všechny nevstupní neurony v síti stejná a je určena aktivační funkcí. Adaptivní dynamika na rozdíl od aktivní dynamiky popisuje počáteční konfiguraci sítě a její změnu v čase. Jednotlivé typy sítí se od sebe liší těmito třemi složkami dynamiky. Dá se tedy říci, že pro specifikaci sítě stačí formulovat jeho organizační, aktivní a adaptivní dynamiku. (Šíma a Neruda, 1996).

Životní cyklus neuronové sítě

Obecně lze životní cyklus neuronové sítě rozdělit do dvou fází (Olej a Hájek, 2010):

- fáze učení,
- fáze života.

Fázi učení můžeme chápat jako adaptaci neuronové sítě. V této fázi se synaptické váhy mění za určitých pravidel neuronové sítě tak, aby zachytily znalosti získané v průběhu učení. Učení se provádí na tzv. **trénovací množině**, což je podmnožina dat popisující problém, který bude neuronová síť řešit. Kvalita trénovací množiny je velice důležitá, protože ovlivní výsledky neuronové sítě. Každá jedna množina vstupů z trénovací množiny a případná množina výstupů se nazývá **vzor**. Trénovací množinou můžeme učit síť vícekrát, v takovém případě každý jeden průchod trénovací množiny nazýváme **epocha** (Olej a Hájek, 2010) nebo jej dle Šíma a Neruda (1996) můžeme nazvat **trénovací cyklus**.

Existují 2 způsoby učení. První je učení s učitelem, který neuronové síti nejenže předkládá vstupy, ale rovněž i správné výstupy z trénovací množiny. Adaptace sítě probíhá v závislosti na tom, jaký výstup síť vygeneruje (správný, nesprávný). Druhý způsob je učení bez učitele, kdy neuronová síť dostává pouze vstupy. Tento způsob učení je založen na schopnosti rozeznat na vstupech stejné nebo podobné vlastnosti, které se snaží slučovat. Neuronová síť se adaptuje na základě výpočtu vzdálenosti mezi vstupy a aktuálními hodnotami. Fáze života následuje po fázi učení. Zde se již synaptické váhy nemění, pouze se využívá naučená síť k řešení nějakého problému, např. klasifikace, predikce (Olej a Hájek, 2010).

Feedforward

Feedforward znamená dopředné šíření signálu v neuronové síti. Jedná se o odezvu neuronové sítě na vstupy excitací neuronů ve výstupní vrstvě. Následuje proces adaptace vah, během

které neuronová síť mění svou konfiguraci např. algoritmem backpropagation. Proces feedforwardu může být popsán slovy následovně (Vondrák, 2009):

1. „Nejprve jsou excitovány na odpovídající úroveň (v rozmezí 0 až 1) neurony vstupní vrstvy.
2. Tyto excitace jsou pomocí vazeb přivedeny k následující vrstvě a upraveny (zesíleny či zeslabeny) pomocí synaptických vah.
3. Každý neuron této vyšší vrstvy provede sumaci upravených signálů od neuronů nižší vrstvy a je excitován na úroveň danou svou aktivační funkcí.
4. Tento proces probíhá přes všechny vnitřní vrstvy až k vrstvě výstupní, kde pak získáme excitační stavy všech jejích neuronů“. (Vondrák, 2009, s. 93).

Backpropagation

Backpropagation je opakem k feedforwardu a znamená zpětné šíření signálu. Adaptace vah probíhá od výstupní vrstvy směrem ke vstupní vrstvě, popisuje Vondrák (2009, s. 94). Hlavním rozdílem je výpočet chyby mezi skutečnou a požadovanou odezvou neuronové sítě. Tato chyba se pak přenáší na nižší vrstvy sítě a upravují se jimi váhy v určitém poměru až k vrstvě vstupní. Cílem je mít chybu co nejmenší.

Dle Šíma a Neruda (1996), nejpoužívanější algoritmus pro adaptivní dynamiku, kterým je backpropagation, má nevýhodu, že používá výpočet gradientu (tečný vektor) pro zjištění minima chybové funkce. Metoda v kombinaci s tímto algoritmem bohužel nachází pouze nejbližší minima chybové funkce. To však nemusí být vždy globální, ale může být jen lokální. Tato skutečnost ovlivňuje výsledky neuronové sítě, protože bylo-li dosaženo pouze lokálního minima (a globální minimum je hodně odlišné), znamená to, že nebyla dostatečně eliminována chyba neuronové sítě a o tuto chybu pak mohou být výsledky nepřesné. Existují způsoby, jak tomu předejít. Tím nejjednodušším řešením, je asi inicializovat váhy náhodně, avšak tak, aby si jednotlivé hodnoty nebyly příliš blízké. Dobrým výsledkům může pomoci taktéž upravený algoritmus backpropagation, který nejenže upravuje váhy, ale taktéž i strmost aktivační funkce, která určuje tzv. míru rozhodnosti jednotlivých neuronů. V neposlední řadě může být dle Volné (2008b) použito evolučních algoritmů pro vytipování možných optim a následně výsledky předány neuronovým sítím, aby dokončily proces optimalizace do globálního minima.

Výsledek také ovlivní vybraná trénovací množina. Je dobré vybrat kvalitní vzory napříč problémovou oblastí. V některých případech je nutné upravit neuronovou síť tím, že například přidáme nebo ubereme skrytou vrstvu či počet neuronů v ní. Ani organizační dynamika nám však nezajistí kvalitní výstupy. Například pokud přidáme více neuronů či vrstev, stoupne výpočetní náročnost a síť je náchylná k tzv. přeučení (overfitting). Přeučení lze vysvětlit tak, že neuronová síť neumí generalizovat, resp. aproximovat výsledky, ale výsledky rovnou kopíruje, protože si dokáže přesně zapamatovat vzory z trénovací množiny. Pro optimalizaci máme k dispozici také koeficient rychlosti učení (learning rate). V případě, že je koeficient malý, neuronová síť pomalu konverguje. Je-li velký, neuronová síť konverguje rychle, ovšem chyba výsledné neuronové sítě roste. Vhodná strategie je začít s nízkým koeficientem a postupně jej zvyšovat. Je zřejmé, že neexistuje pravidlo, jakým neuronovou síť nastavit správně. Nelze se však čemu divit, kdyby existovalo ono pravidlo, pak by šlo popsat klasickým algoritmem a nepoužívalo by se jedné z metod umělé inteligence. (Šíma a Neruda, 1996).

2.3 Perceptronová síť

Perceptronová síť je jednou z nejpoužívanějších modelů neuronové sítě. Základním elementem je perceptron vynalezený v roce 1957 Frankem Rosenblattem. Může existovat více druhů perceptronové sítě v závislosti na její organizační, aktivní a adaptivní dynamice. Rovněž tyto sítě mohou být označovány jako vícevrstvý perceptron, jak popisuje Berka (2008). Nejčastěji se používá v kombinaci s aktivační funkcí sigmoida a s algoritmem backpropagation (Šíma a Neruda, 1996). Tato specifikace je popsána níže.

„Činnost perceptron má jednoduchou geometrickou interpretaci. Vstupní podněty mohou představovat hodnoty vstupních atributů nějakého objektu (např. teplota, výška, váha,...). Každý objekt lze pak reprezentovat jako bod $x = x_1, x_2, \dots, x_m$ v m -rozměrném prostoru. Bod x leží v jedné ze dvou částí prostoru, oddělených od sebe rozděľující nadrovinou. Body, ležící v jedné části prostoru, můžeme považovat za obrazy objektů, patřících do téže třídy. Perceptron lze tedy považovat za (globální) lineární klasifikátor objektů do dvou tříd. Pro lineárně separabilní třídy je vhodné použít perceptron.“ (Berka, 2008, s. 120).

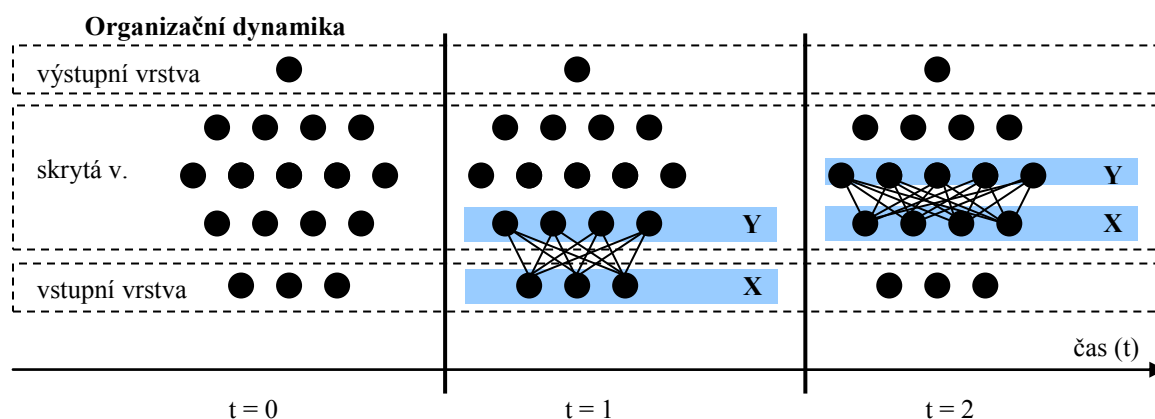
2.3.1 Organizační dynamika

Perceptronová síť se skládá z 1 vstupní vrstvy, několika skrytých vrstev a z 1 výstupní vrstvy, z čehož vyplývá, že minimálně je tvořena 3 vrstvami. A jak tvrdí Vondrák (2009, s. 93), předpokladem je „tzv. úplné propojení neuronů, tedy každý neuron nižší vrstvy je spojen se všemi neurony vrstvy vyšší“.

„Množinu n vstupních neuronů označíme X a množinu m výstupních neuronů Y . Neurony značíme indexy i, j apod. a ζ_j představuje reálný vnitřní potenciál a y_j reálný stav, resp. výstup neuronu j . Spoj od neuronu i k nevstupnímu neuronu j je ohodnocen reálnou synaptickou váhou w_{ji} a $w_{j0} = \text{bias}$ “. Dále $j \leftarrow$ označuje množinu všech neuronů, z nichž vede spoj do neuronu j a které jsou tedy vstupem neuronu j , a $j \rightarrow$ je množina neuronů, do nichž vede spoj z neuronu j a kterým je proto neuron j vstupem.“ (Šíma a Neruda, 1996, s. 52).

Je nutné zde poznamenat, že množiny X a Y jsou chápány jako relativní ve vztahu k organizační struktuře, nejedná se tedy pouze o neurony vstupní vrstvy (X) a neurony výstupní vrstvy (Y). Níže uvedené vzorce pracují vždy se sousedními neurony dvou vrstev, mezi kterými existuje synapse, kterým je určena váha w_{ji} . Neurony se obecně značí jako kolečka a vazby mezi nimi jsou čáry mezi kolečky. Vstupní vrstva bývá dole anebo vlevo v závislosti na orientaci grafu.

Obr. 2.3 – Vztah množin X a Y k organizační dynamice



Zdroj: vlastní zpracování

2.3.2 Aktivní dynamika

Aktivní dynamika vypočítává stavy všech neuronů postupně dle času t . Pro výpočet hodnoty vnitřních potenciálů, které v tomto případě mohou nabývat reálných hodnot, je zapotřebí jen dvou vrstev sousedních neuronů, resp. pro ty vrstvy neuronů j , které mají známé vstupy. Postupuje se postupně od první skryté vrstvy k výstupní vrstvě. Dá se též říci, že „v čase t jsou aktualizovány neurony v t -té vrstvě. Z vnitřního potenciálu je pak stanoven reálný stav neuronu j pomocí diferencovatelné aktivační funkce (standardní sigmoidy), která spojitě aproximuje ostrou nelinearitu.“ (Šíma a Neruda, 1996, s. 53).

$$\xi_j = \sum_{i \in j_{\leftarrow}} w_{ji} \cdot y_i \quad (2.2)$$

Vzorec 2.2 popisuje vnitřní potenciál j -tého neuronu, kde (Šíma a Neruda, 1996, s. 53):

y_i stav neuronu, kde $i \in X$.

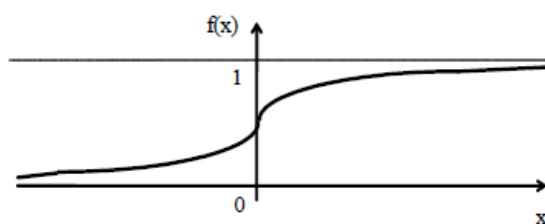
Jako aktivační funkce se používá standardní sigmoida. Funkce sigmoidy je plynule diferencovatelná, což je podstatné pro výpočet gradientů v algoritmu backpropagation. (Šíma a Neruda, 1996):

$$\sigma(\xi) = \frac{1}{1 + e^{-\lambda \cdot \xi}} \quad (2.3)$$

kde:

λ strmost sigmoidy (ve standardních modelech platí, že $\lambda = 1$).

Obr. 2.4 – Sigmoida - průběh funkce



Zdroj: Volná (2008a)

2.3.3 Adaptivní dynamika

Algoritmus backpropagation se snaží minimalizovat chybu sítě. Ta se nejprve vypočte pro jednotlivé vzory z trénovací množiny (viz. rovnice 2.4). Suma těchto odchylek je pak chyba celé sítě (viz. rovnice 2.5). (Šíma a Neruda, 1996, s. 51).

$$E_k(w) = \frac{1}{2} \sum_{j \in Y} (y_j(w, x_k) - d_{kj})^2 \quad (2.4)$$

kde:

$E_k(w)$ chyba vzoru z trénovací množiny,
 $y_j(w, x_k)$ výstup j neuronu při váhách a vstupech k vzoru,
 d_{kj} požadovaný výstup pro vstup k vzoru pro j neuron.

Vzorec 2.4 je ve své podstatě metrika eukleidovské vzdálenosti, která počítá vzdálenost mezi správným výsledkem a výsledkem, který spočítala neuronová síť.

$$E(w) = \sum_{k=1}^p E_k(w) \quad (2.5)$$

kde:

$E(w)$ chyba celé sítě,
 k vzor z trénovací množiny,
 p počet vzorů.

Jak uvádí Šíma a Neruda (1996, s. 54) „nová konfigurace $w^{(t)}$ v čase $t > 0$ se vypočte:“

$$w_{ji}^{(t)} = w_{ji}^{(t-1)} + \Delta w_{ji}^{(t)} \quad (2.6)$$

kde:

t čas
 w_{ji} váha spoje od neuronu i k nevstupnímu neuronu j .

Šíma a Neruda (1996, s. 54) pokračují. „Změna vah $\Delta w^{(t)}$ v čase $t > 0$ je úměrná zápornému gradientu chybové funkce $E(w)$ v bodě $w^{(t-1)}$:“

$$\Delta w_{ji}^{(t)} = -\varepsilon \cdot \frac{\partial E}{\partial w_{ji}} \cdot (w^{(t-1)}) \quad (2.7)$$

kde:

ε koeficient rychlosti učení (learning rate).

Samotný výpočet parciální derivace je příliš obsáhlý, a proto zde nebude uveden. Pro potřeby implementace je třeba však použít jeho výsledku viz. vzorec 2.8. Do tohoto vzorce se za $\frac{\partial E_k}{\partial y_j}$ dosazují výsledky rovnic 2.9, kterou se vypočítává dle Vondrák (2009) „rozdíl mezi skutečnou a požadovanou odezvou neuronu j vzoru k vnější vrstvy“ (tj. horní vrstva) a rovnice 2.10, která vypočítává to samé, ovšem jen pro vnitřní vrstvy (tj. nižší vrstvy). Dále jsou uvedeny ony vzorce, jak je uvádí Šíma a Neruda (1996, s. 56, 57):

$$\frac{\partial E_k}{\partial w_{ji}} = \frac{\partial E_k}{\partial y_j} \cdot \lambda_j \cdot y_j \cdot (1 - y_j) \cdot y_i \quad (2.8)$$

kde:

- λ strmost sigmoidu,
- y_j stav neuronu, kde $j \in Y$,
- y_i stav neuronu, kde $i \in X$,

$$\frac{\partial E_k}{\partial y_j} = y_j - d_{kj} \quad (2.9)$$

kde:

- d_{kj} požadovaná hodnota výstupu pro vzor k .

Pro nižší vrstvy se provede následující výpočet:

$$\frac{\partial E_k}{\partial y_j} = \sum_{r \in J \rightarrow} \frac{\partial E_k}{\partial y_r} \cdot \lambda_r \cdot y_r \cdot (1 - y_r) \cdot w_{rj} \quad (2.10)$$

Algoritmus backpropagation nejdříve inicializuje neuronovou síť tak, že nastavení náhodné váhy synapsím reálným číslem a chyby synapsí nastaví na nulu. Poté začíná učení sítě formou předkládání jednotlivých vzorů z trénovací množiny. Probíhá feedforward resp. dopředné šíření signálu a výpočet aktivačních funkcí jednotlivých neuronů. Následuje výpočet parciálních derivací chybové funkce postupně od horních vrstev k nižším. Vypočítává se gradient parciální chybové funkce, který ovlivní jednotlivé váhy synapsí. Na závěr se spočte chyba celé sítě a v případě, že je malá, učení se ukončí. Není-li chyba dostatečně malá,

pokračuje se předložením dalšího vzoru a celý proces se opakuje, kromě počáteční inicializace neuronové sítě.

2.4 Kohonenova samoorganizující se mapa

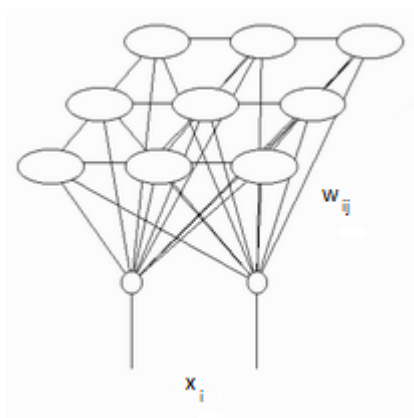
Tento další druh neuronové sítě se řadí k architektuře vycházející ze soutěžního typu. Sít' popsal fin Teuvo Kohonen v 80. letech 20. století a navázal tak na princip soutěžení, který navrhl von der Malsburg. Oproti Perceptronové síti je Kohonenova samoorganizující se mapa novější a řeší jiný typ úloh. Řadí se mezi sítě bez učitele, avšak existují i její modifikace, které ji mění na síť s učitelem. Kohonenova samoorganizující se mapa umí aproximovat vstupní hodnoty na základě vektorové kvantifikace. Tento druh sítě se používá pro shlukovou analýzu. Existuje více modifikací této sítě. Níže je popsán pouze jeden typ určité modifikace. (Šíma a Neruda, 1996)

2.4.1 Organizační dynamika

Topologie sítě je dvouvrstvá. Neurony vstupní vrstvy, též kompetiční vrstva, jsou propojeny se všemi neurony výstupní vrstvy. Neurony ve výstupní vrstvě mohou být taktéž mezi sebou propojeny. Výstupní vrstva se nejčastěji zobrazuje ve dvojdimenzionálním prostoru, jak je zobrazeno na obrázku níže. Vítězný neuron z výstupní vrstvy může být pouze jeden, někdy též nazýván Grandmother cell. (Vondrák, 2009).

„Neurony v kompetiční vrstvě plní funkci reprezentantů a jsou uspořádány do struktury (nejčastěji dvojrozměrné mřížky). Ta určuje, které neurony spolu v neuronové síti sousedí. Synaptické váhy každého neuronu představují souřadnice udávající konkrétní polohu neuronu v prostoru. Základní princip učení Kohonenovy samoorganizující se mapy spočívá ve stanovení vzdálenosti d_j mezi předkládanými vstupními vektory x_i a vektory souřadnic (synaptické váhy) neuronů w_{ji} umístěných v kompetiční vrstvě.“ (Olej a Hájek, 2010, s. 62, 63).

Obr. 2.5 – Kohonenova samoorganizující se mapa



Zdroj: http://automatizace.hw.cz/images/kohonen_maps/smallsitova_struktura2.gif

2.4.2 Aktivní dynamika

Jelikož v této síti je výstupem pouze jeden neuron, má aktivní dynamika jiný tvar než např. u perceptronové sítě. Nejedná se o žádnou ostrou linearitu, apod. Pouze jeden vítězný neuron ve výstupní vrstvě může mít hodnotu 1, zatímco všechny ostatní mají hodnotu 0. Neurony nemají práh. Vítězný neuron je ten, jehož vektor je nejbližší k vstupnímu vektoru, resp. d_j z rovnice 2.11 je minimální. (Vondrák, 2008).

„Výstup tohoto neuronu je aktivní, výstupy ostatních neuronů jsou neaktivní. Okolo vítěze se dále vytvoří okolí, do kterého se zahrnou ty neurony, které se podle zvoleného kritéria nejvíce podobají vítězi. Váhy těchto neuronů se pak modifikují v procesu učení.“ (Olej a Hájek, 2010, s. 63).

$$d_j = \sum_{n=1}^n (x_i(t) - w_{ij}(t))^2 \quad (2.11)$$

kde:

d_j stav neuronu, kde $j \in Y$,

x_i stav neuronu, kde $i \in X$,

t čas,

w_{ji} váha spoje od neuronu i k nevstupnímu neuronu j .

2.4.3 Adaptivní dynamika

Adaptivní dynamika je popsána v následující rovnici. Její princip vychází z tzv. laterální inhibice, která posiluje vždy vítězný neuron a ostatním váhy nemění, anebo jim změnu váhy negativním směrem. Důležitým elementem je koeficient učení, který je doporučeno měnit v čase. Jak tvrdí Šíma a Neruda (1996), na začátku učení by jeho hodnota měla být blízko jedné, s postupem času by se měl koeficient snižovat.

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t) \cdot (x_i(t) - w_{ij}(t)) \quad (2.12)$$

kde:

η koeficient učení.

2.5 Sít' MADALINE

Sít' MADALINE je zkratkou pro Multiple ADALINE. A tak jako v Perceptronové síti tvoří síť perceptron, zde síť tvoří neurony zvané ADALINE (ADaptive LINEar Element). Princip ADALINE neuronu je podobný perceptronu. Tyto neurony byly navrženy Wodrowem a Hoffem v 50. letech 20. století. Sít' je schopna najít závislosti mezi vzory a separovat je. (Šíma a Neruda, 1996)

2.5.1 Organizační dynamika

„Organizační dynamika MADALINE a značení parametrů sítě je stejná jako u sítě perceptron, avšak místo perceptron uvažujeme ADALINE“. (Šíma a Neruda, 1996, s. 64). Někteří autoři jako Volná (2008a) uvádějí organizační dynamiku jako třívrstvou síť, kde všechny neurony ve skryté vrstvě a ve výstupní vrstvě jsou tvořeny neurony typu ADALINE.

2.5.2 Aktivní dynamika

Jak uvádí Šíma a Neruda (1996), výstupy sítě MADALINE mohou být reálná čísla. Neuron ADALINE realizuje lineární funkci:

$$y_j = \sum_{i=0}^n w_{ji} \cdot x_i \quad (2.13)$$

kde:

- y_j stav neuronu, kde $j \in Y$,
- w_{ji} váhy mezi vstupním a výstupním neuronem
- x_i stav neuronu, kde $i \in X$,
- w_{ji} váha spoje od neuronu i k výstupnímu neuronu j .

2.5.3 Adaptivní dynamika

Síť MADALINE patří do kategorie sítí s učitelem. Vstupy mohou být reálná čísla, stejně jako výstupy. Každému tréninkovému vzoru je přiřazen správný výstup. Odchylka mezi vypočteným výstupem sítě a požadovaným výstupem je značena jako chyba. Adaptivní dynamika má za cíl minimalizovat tuto chybu. Změna vah se provádí po průchodu každého tréninkového vzoru dle vzorce 2.14. (Šíma a Neruda, 1996).

$$w_{ji}(t+1) = w_{ji}(t) - \varepsilon \cdot x_{ki} \cdot (y_j(t) - d_{kj}) \quad (2.14)$$

kde:

- ε koeficient učení,
- k tréninkový vzor.

3. Analýza problematiky, specifikace procesů a datových toků

Neuronové sítě umí řešit široké spektrum úkolů. Tato diplomová práce se zabývá konkrétním úkolem, kterým je aplikace umělé inteligence poskytující relevantní informaci podle profilu uživatele. Je známo, že metody umělé inteligence potřebují mnoho dat, aby mohly být efektivně využity. Diplomová práce bude pracovat nad reálnými daty z informačního systému, který byl vytvořen v rámci bakalářské práce viz. Zulák (2008) a který se zabývá zpracováním zpravodajských článků. Tento informační systém, který je zevrubně popsán v kapitole 3.2, není předmětem diplomové práce, pro řešený model poskytuje pouze data. Nejprve však následuje podkapitola popisující problematiku zpracovaných dat obecně.

3.1 Analýza problematiky

Navržená neuronová síť bude pracovat nad zpravodajskými daty, resp. články z on-line novin, časopisů, blogů apod. Konfigurace sítě bude nastavena na základě profilu uživatele. Jednotlivé použité atributy budou specifikovány později, nicméně mezi takovéto atributy lze zařadit pohlaví, věk, region (kraj), který uživatele zajímá, apod. Podle úlohy, kterou by měla neuronová síť řešit, bude vybrána vhodná topologie sítě. Podrobněji se neuronovou sítí zabývá čtvrtá kapitola.

Existující informační systém sbírá periodicky nové články ze zpravodajských serverů a ukládá je do databáze v procesu aktualizace, který je popsán v navazující podkapitole. Existuje zde více faktorů, které ovlivňují rozsah i kvalitu ukládaných článků. Ve svém důsledku se ukládá libovolná z kombinací, přičemž titulek článku a datum vydání je ukládán vždy:

- titulek článku,
- datum vydání,
- abstrakt článku,
- odkaz na přílohu (jednu nebo několik), která, je-li v podobě obrázku, je u zprávy zobrazena,
- odkaz na komentáře ke článku.

Obr. 3.1 – Příklad zobrazení dat s maximem informací



Zdroj: www.newsservis.com

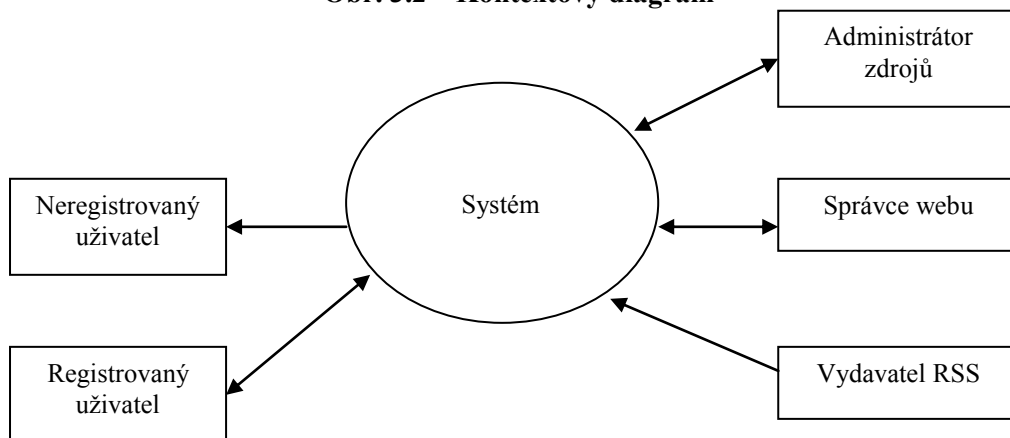
Na obr. 3.1 je vidět článek s maximem uložených informací. U tohoto článku byl uložen titulek, datum vydání, abstrakt, příloha (obrázek) a odkaz na komentáře. Mimo jiné je zde i název zdroje, ze kterého zpráva pochází a statistika počtu kliknutí. Články se vždy váží ke svému zdroji, resp. serveru, ze kterého pochází. Zdroje jsou zařazeny do 7 hlavních kategorií (zpravodajství, sport, bulvár, aj.) a případně i podkategorií, tak jako bývá zvykem u jiných internetových stránek tohoto typu.

Registrovaný uživatel si může oblíbené zdroje ukládat a sledovat jejich obsah. Zdroje mu jsou ihned k dispozici a nemusí je znovu vyhledávat, což šetří čas. Existuje ještě řada dalších funkcí IS, které zefektivňují práci uživatele a podílí se na tzv. personalizaci stránek. Tento druh personalizace, který je založen obecně na uložení nějakého parametru a jeho následného přečtení, je však poměrně omezený. Příkladem může být právě již zmiňované uložení oblíbených zdrojů. U tohoto typu personalizace je třeba, aby uživatel vždy učinil nějakou akci, např. kliknul na tlačítko uložit. Neuronové sítě sice také potřebují určitou konfiguraci, ale jenom ze začátku. V průběhu času se vyvíjí samy na základě podnětů, umějí reagovat, přizpůsobit se, aproximovat, umí se naučit vzor určitého chování, apod. Nabízí se zde možnost využít metod neuronových sítí pro řešení úloh, které přispějí k lepší personalizaci informačního systému a k lepší spokojenosti uživatele.

3.2 Specifikace stávajícího informačního systému

Data, se kterými bude neuronová síť pracovat, jsou z informačního systému, který byl implementován na základě bakalářské práce viz. Zulák (2008). Informační systém, který bývá v tomto kontextu označován též jako on-line agregátor zpráv, je umístěn na internetu a zpracovává články z jednotlivých zpravodajských serverů (zdrojů). Uživatelům zobrazuje na jednom místě titulky a abstrakty článků (u některých i přílohy jako jsou fotografie, či záznam zvuku) z mnoha různorodých zdrojů (včetně zahraničních) a poskytuje funkce, z nichž nejvíce používanou je vyhledávání. Pokud si chce uživatel přečíst celý článek, je po kliknutí na odkaz přesměrován na internetovou stránku s celým obsahem, která je umístěna na zpravodajském serveru, ze kterého byl článek původně načten. Tento druh IS šetří uživatelům zejména čas, který by strávili navštěvováním jednotlivých zdrojů. Dále pak nabízí jednotné ovládání, vyhledávání ve více zpravodajských zdrojích najednou včetně rozšířeného vyhledávání, apod. Způsob, jakým systém pracuje a jednotlivé jeho procesy, popisuje tato kapitola.

Obr. 3.2 – Kontextový diagram



Zdroj: Zulák (2008, s. 12)

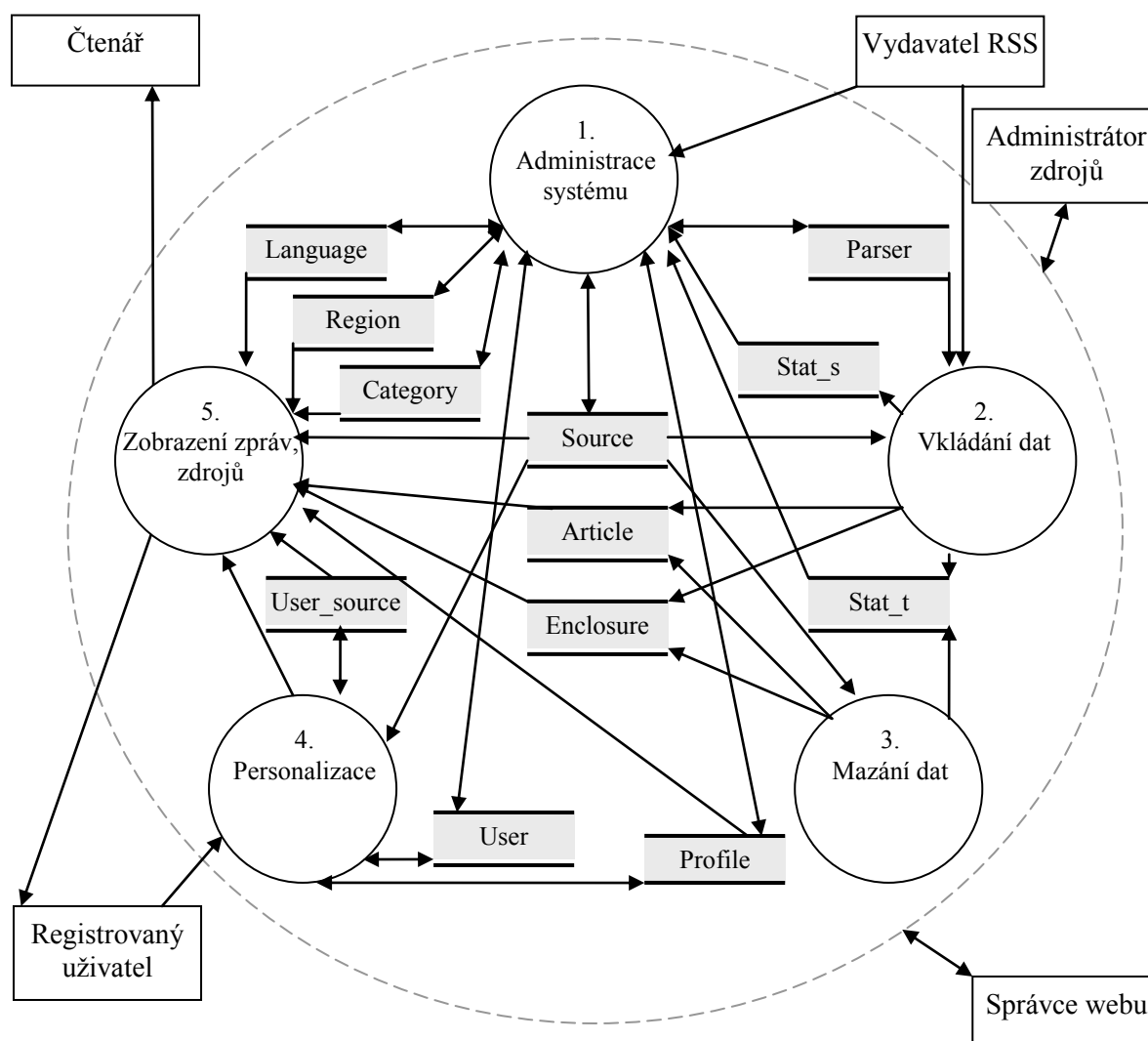
Se systémem pracují uživatelé, kteří jsou rozděleni do dvou skupin: neregistrovaný a registrovaný uživatel. Neregistrovaný uživatel data pouze přijímá, má jen omezenou možnost jak výstupní sestavy měnit. Registrovaný uživatel má v IS svůj profil a nastavené parametry, na základě kterých jsou mu výstupní sestavy předgenerovány. O správu IS se stará správce webu, který má plný přístup do administrační zóny a administrátor zdrojů, který spravuje pouze zdroje. Vydavatel RSS v kontextovém diagramu představuje onen zdroj, resp. subjekt

(internetové stránky novin, zpravodajský server, aj.), který disponuje zdroji dat ve formátu RSS.

Informační systém používá jako zdroj dat výhradně formát RSS, ačkoliv zpracování jiného formátu dat lze snadno přidat. RSS (Really Simple Syndication) je souborový formát používaný pro syndikaci obsahu, což je moderní metoda pro sdílení rychle měnících se informací na internetu. Kromě RSS, který je podmnožinou XML, se můžeme na internetu setkat s dalšími formáty, kterými jsou např. Atom, CVS, CFD, aj. RSS je však v současnosti nejrozšířenějším formátem zejména na českých zpravodajských serverech.

Informační systém je rozdělen do 5 hlavních procesů. První 3 procesy slouží ke správě IS, zbylé dva pak prezentují přidanou hodnotu systému, která spočívá zejména v různých parametrizovaných výstupních sestavách. DFD je zakreslen dle notace DeMarco a Yourdon, kterou popisují Dennis, Wixom a Roth (2009, s. 175).

Obr. 3.3 – DFD 0. úrovně



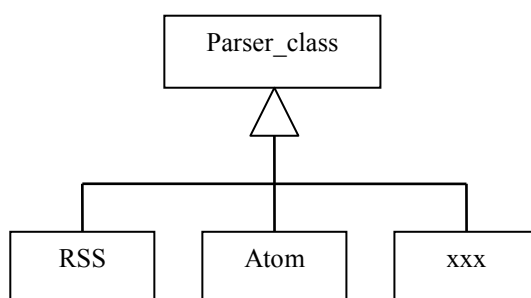
Zdroj: upraveno dle Zulák (2008, s.15)

Administrace systému

Do administrace systému má plný přístup správce webu a omezený správce zdrojů. Tento proces obsahuje funkce pro řízení IS. Správce webu může spravovat jazyky, regiony, kategorie, tzv. parsery (skript, který zpracovává obsah zdroje), statistiky a v neposlední řadě i zdroje příp. jejich články s přílohami. Správce zdrojů může spravovat pouze zdroje a jim příslušné články s přílohami.

Každý zdroj v systému je jedinečný a obsahuje zejména URL odkaz na externí zdroj dat a parser, kterým se data zpracují. V současnosti se používá formát RSS. Informační systém lze rozšířit, aby uměl zpracovat i jiné souborové formáty jak je znázorněno na obr. 3.4.

Obr. 3.4 – UML diagram - parser



Zdroj: Zulák (2008, s. 21)

Vkládání dat

Pod vkládáním dat se rozumí načtení a zpracování zdroje. To probíhá automatizovaně periodicky anebo tento proces může být spuštěn manuálně z administrativní zóny. Perioda aktualizace může být u každého zdroje nastavena jinak v závislosti na různých parametrech. Informační systém obsahuje speciální algoritmus, určující další pravidla pro aktualizaci, pravidla pro tzv. odstavení zdroje v případě, že se ho nepodařilo načíst správně, automatické oznámení chyb správci zdrojů, apod.

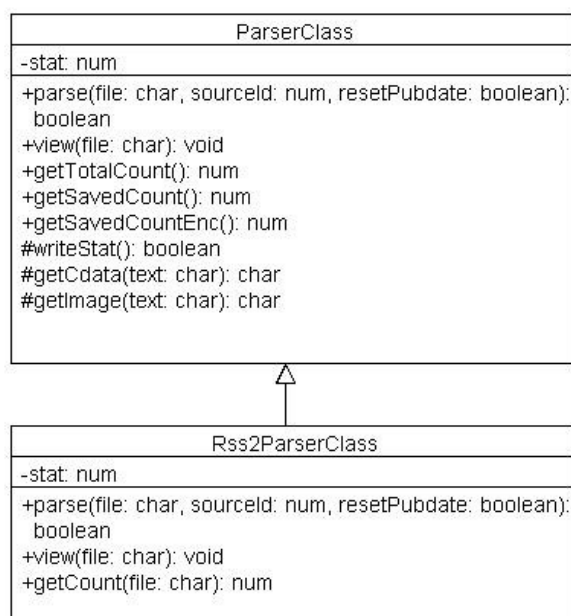
Práce se zdroji je v současnosti založena na principu push. Jak uvádí Vrána (2010), do popředí se dostává i opačný princip pull, který má opačný tok informací. Zástupcem takového principu může být protokol PubSubHubbub, kde vydavatel obsahu komunikuje s prostředníky a ti posílají data konečným odběratelům. V tomto případě tento IS by byl onen prostředník. Protokol je založen na notifikaci, která přichází od vydavatele obsahu k prostředníkům. Prostředník na základě obdržené notifikace načte nový obsah. Odpadá zde nutnost periodické aktualizace a snižuje se počet přenesených dat, což má vliv i na rychlost informačního systému. Využít výhod principu pull je cíl, tohoto dílčího procesu.

Důležitou částí v procesu vkládání dat je tzv. parser (obr. 3.5), který zpracovává načtená data tak, aby se dala informačním systémem využít. Každý zdroj může mít jiný parser, čím se zvyšuje variabilita systému. Současný parser pro RSS dokáže zpracovat formáty verze 1.0 a 2.0. Umí zpracovat i přílohy k článkům a zařadit je dle typu (obrázek, video, aj.) a to i v případě, že RSS není validní. Parser prochází neustálým vývojem a jeho funkcionality je rozšiřována tak, aby obstál při zpracování různých anomálií, jakým je bezesporu i nastavení data vydání jednotlivých článků. Některé zdroje nejenže tento atribut zcela vynechávají, jiné

ho zapisují ve špatném formátu. Jsou i zdroje, které datum vydání článků dokonce posunují do budoucnosti, přičemž článek je dostupný již v době zpracování parserem.

Zpracovaná data jsou následně uložena v databázi. S ohledem na rozšiřitelnost informačního systému jsou data v databázi uložena ve znakové sadě UTF-8, která dovoluje uložit i jiné znaky jako např. azbuku. Ve výsledku se mohou na jedné HTML stránce zobrazit česky i rusky psané články a budou čitelné. Jakmile jsou data uložena v databázi, spojení se zdrojem se ukončuje až do doby jeho příští jeho aktualizace.

Obr. 3.5 – Diagram tříd



Zdroj: vlastní zpracování

Metoda *parse* zpracovává zdroj a ukládá jeho obsah do databáze. Metoda *view* pouze zpracuje obsah a bez ukládání do databáze jej zobrazí. Tato metoda je používána v administrativní zóně, aby se zjistilo, zda parser vůbec dokáže zdroj načíst, apod. Metoda *getCount* vrací počet článků ve zdroji. Metody *getCdata* a *getImage* pracují s jednotlivými články a provádí různé úpravy textu. Zbylé metody slouží ke statistickým účelům.

Mazání dat

Mazání dat odstraňuje zejména staré články a přílohy z databáze. Tento proces je plně automatizovaný. Každý zdroj má nastavenou dobu archivace ve dnech, po kterou články a přílohy článku zůstávají v databázi. Po překročení této doby jsou patřičná data smazána.

V tomto procesu se také resetuje potřebná statistika (počet kliknutí na článek za den, apod.), mažou se registrovaní uživatelé, kteří dosud nepotvrdili svou registraci přes autorizační klíč a provádí se další potřebné kroky, které mají za cíl odstranění dat z databáze.

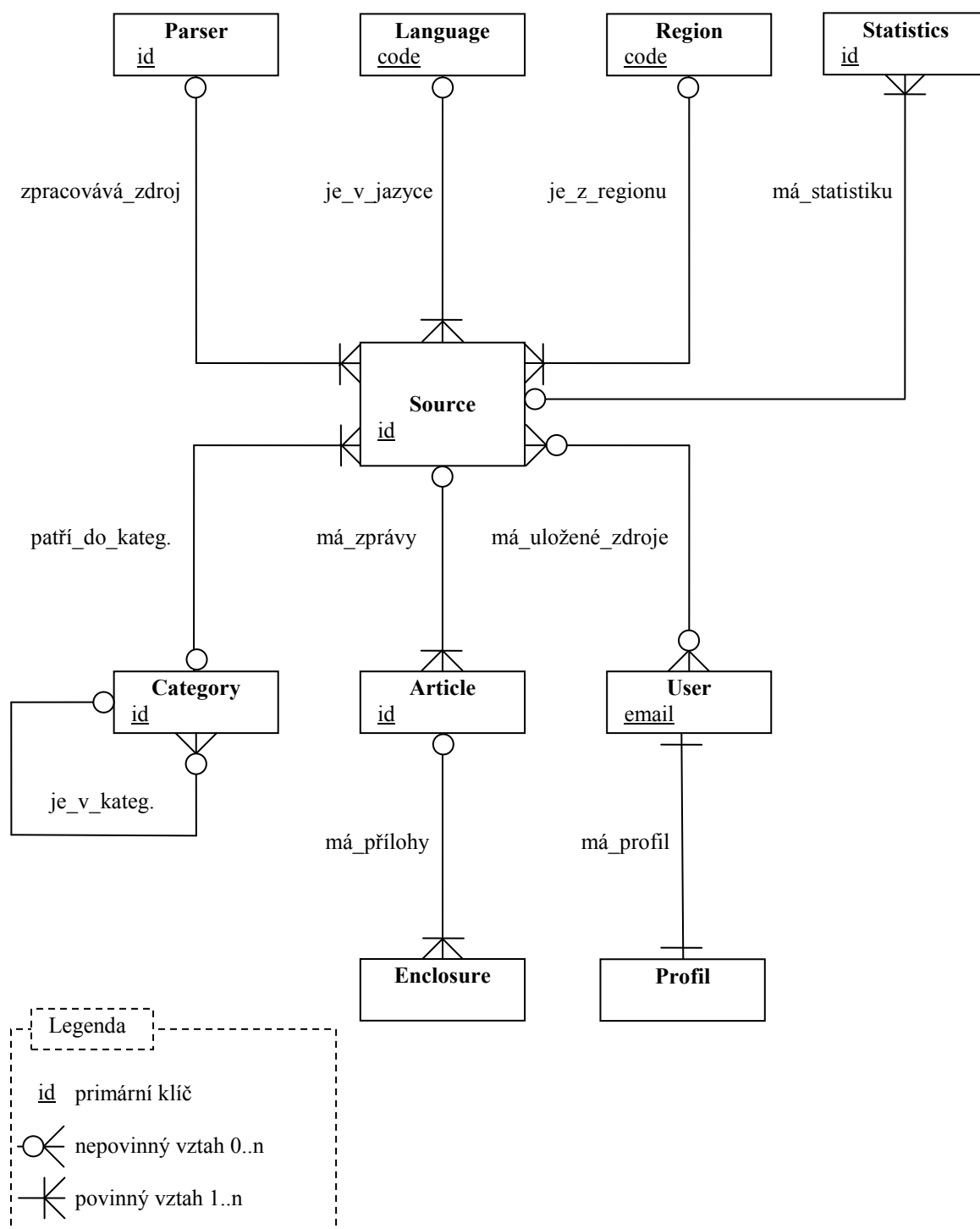
Personalizace

V kontextovém diagramu (obr. 3.2) je vidět, že uživatelé, kteří používají IS ke svému prospěchu, jsou dvojího typu. Může jim být neregistrovaný a registrovaný uživatel. Proces personalizace se týká pouze registrovaných uživatelů. Registrovaný uživatel může používat více funkcí IS než neregistrovaný čtenář. Proces personalizace se však nezabývá zobrazováním článků na míru uživateli, nýbrž agendou s registrovaným uživatelem spojenou. Jedná se zejména o samotnou registraci nových uživatelů (ověřování e-mailů), změnu hesla, nastavení uživatelského profilu, včetně uložení nastavení prostředí a nastavení parametrů pro výstupní sestavy. Proces personalizace tedy předchází proces Zobrazení článků, zdrojů. Pokud si registrovaný uživatel chce zobrazit sestavy, pak se nejprve načte jeho profil, parametry pro výstup a ty jsou předány procesu Zobrazení článků, zdrojů, který výslednou sestavu vygeneruje na základě těchto parametrů. Příkladem může být např. zobrazení jen článků z určitých zdrojů, které uživatel preferuje a uložil si je.

Zobrazení

Proces zobrazení obsahuje většinu funkcí informačního systému. Zde se provádí veškeré filtrování, vyhledávání, generování výstupních sestav, řazení výsledů, stejně jako generování přehledů (právě čtené články, přehled nových článků za kategorii, aj.). Dále pak zobrazení statistik například čtivost zdroje, čtivost článku za den, celková statistika čtivosti článku. Zobrazení příloh a zobrazení zdrojů jsou zde také řešeny. Výsledky jsou dále generovány na základě tzv. šablony zobrazení (články zdroje v tabulce, články zdroje v řádcích, chronologický výpis s vazbou na zdroj s maximem informací, chronologický výpis bez vazby na zdroj). Některé funkcionality jsou řešeny v aplikační vrstvě IS, jiné jsou řešeny v prezenční vrstvě.

Obr. 3.6 – Konceptuální datový model (ERD)



Zdroj: upraveno dle Zulák (2008, s. 14)

Z ERD je vidět, že nejdůležitější entitou je *Source*, resp. zdroj. *Source* musí mít přiřazen právě jeden *parser*, *language*, *region* a *category*. *Category* může být ve vztahu sama k sobě,

tvoří tak stromovou strukturu. *Source* může mít více *article* a může figurovat ve více *statistics*. *User* může mít uložen více *source* a jeden *source* může mít uloženo více *user*. Tento vztah M:N je nahrazen vazební tabulkou *user_source*, která obsahuje uložené zdroje uživatele. Posledním vztahem je vztah mezi *article* a *enclosure*, který je 1:N. *Article* může mít nula až nekonečně mnoho *enclosure*. Při sestavení diagramu byla použita notace dle Conolly, Begg a Holowczak (2009, s. 489).

U entit Language a Region je použit jako primární klíč kód. Tento kód se skládá z následujícího:

- Entita Language: ISO 639-1 (jazyk je identifikován dvěma písmeny), např.:

code:
CS
 ↓
 čeština

- Entita Region: NUTS0 a NUTS3 (první slouží pro stát a druhý pro kraje), např.:

code:
CZ
 ↓
 Česká republika

code:
CZ 080
 ↓ ↘
 Česká republika Moravskoslezský kraj

4. Systémový návrh modelu

Nejprve je v této kapitole poukázáno na rozdílné vlastnosti jednotlivých typů neuronové sítě a popsány úkoly, které by síť mohly řešit. Následně je vybrán a prezentován výsledný model neuronové sítě ve vztahu na informační systém.

Na začátku je třeba vysvětlit, co přesně znamená pojem relevantní informace. V diplomové práci je tento pojem chápán podobně jako dle slovník-cizich-slov.cz (2012), který tvrdí, že relevantní informace je „poskytovaná informace, podle potřeby jedince ve specifické situaci“.

V kontextu výše popsaného informačního systému může být chápána relevance informace z mnoha pohledů. Následující výčet uvádí příklady, jež některé ani nevyžadují podporu neuronových sítí:

- Výsledky vyhledávání řazené dle potřeby jedince (počet vyhledaných slov, řazení dle času vydání, aj.).
- Uživatel sleduje téma a jsou mu zobrazovány zdroje, které se tímto tématem nejvíce zabývají, resp. o něm píší.
- Zobrazení článků v závislosti na geolokaci uživatele (např. je-li uživatel z Ostravy, zobrazují se mu přednostně články týkající se Ostravska).
- IS slučuje články do témat, kterých se týkají např. mistrovství světa ve fotbale, uživateli se pak při sledování daného tématu zobrazí relevantní články.
- Uživatele zajímají určitá témata (zájmy) a jsou mu zobrazovány články, obsahující tato témata na základě různé síly preference.

V informačním systému existují omezení, která některá řešení diskvalifikují. Významným nedostatkem IS je, že zatím nedisponuje žádným mechanismem pro rozbor a zpracování textů článků. Pokud by byla například uvažována možnost geolokace zpráv, vyžadovalo by to, aby měl každý článek specifikován svou regionální příslušností. IS sice uchovává regionální příslušnost zdrojů, bohužel však ne příslušnost článků. A protože platí, že zdroj, který má sídlo např. v Českých Budějovicích může vydávat články o Ostravsku, tato klasifikace by nebyla relevantní. Zbytek diplomové práce proto popisuje poslední příklad, kdy jsou uživatelům zobrazovány články týkající se témat, které ho zajímají. Toto řešení bylo vybráno s ohledem

na vlastnosti a omezení IS, budoucí směr růstu IS i možnosti praktické realizace neuronové sítě.

4.1 Výběr sítě

Úkol, který má neuronová síť řešit, je následující: „Uživatelé zajímají určitá témata (zájmy) a jsou mu zobrazovány články, obsahující tato témata na základě různé síly preference“. Z toho vyplývá, že je třeba pracovat s následujícími komponentami:

- čtenář,
- články,
- témata,
- velikost zájmu o téma.

Pro lepší přehlednost je v těchto kapitolách uveden pojem *čtenář* a nikoliv uživatel. Čtenářem může být teoreticky registrovaný i neregistrovaný uživatel, který pracuje s články. V následujícím textu je čtenářem pouze registrovaný uživatel, o kterém jsou v IS evidována data a hlavně jeho zpětná vazba, která je potřebná pro učení sítě, je jednoznačně identifikovatelná. Tentýž model sítě lze aplikovat i na ony neregistrované uživatele, tam by však bylo třeba vyřešit identifikace jiným způsobem.

Nejprve je třeba klasifikovat úlohu a přiřadit jí typ sítě. Vstupů je libovolně mnoho a představují čtenáře. Článků je taky libovolně mnoho a počet článků je větší než počet čtenářů. Tyto komponenty nelze odhadnout, ani dopředu vědět jejich počet. Zatímco omezení počtu čtenářů je problém, omezit počet článků lze, přičemž ale tímto omezením klesá efektivita sítě. Čtenáři se mohou zajímat o různorodá témata, kterých může být taktéž libovolně mnoho, ovšem dá se předpokládat, že se témata u různých čtenářů budou opakovat. Témata lze případně i omezit, resp. nastavit napevno množinu témat, ze které si uživatelé mohou vybrat. Čtenář bude mít na výběr z více témat a bude moci jim nastavit různou velikost zájmu, čímž vyjádří různé preference.

Výše zmíněné komponenty musí mezi sebou vzájemně spolupracovat tak, aby na začátku byl čtenář, žádající relevantní články. Tudíž výstup sítě musí být doporučované články. Otázkou volby konkrétního typu sítě se zabírají další odstavce.

Teoretická část diplomové práce popisuje Kohonenovu samoorganizující se mapu jako typ sítě, který je často používán pro jeho schopnost slučovat objekty. Tato síť zde uplatnění však nenajde, neboť tento typ sítě má na výstupu vždy jen jeden neuron aktivní, což je zde omezující. Tento neuron sice může reprezentovat skupinu objektů, která by mohla představovat například články týkající se nějakého tématu, přičemž ono téma by charakterizoval výstupní neuron. Nicméně pak by bylo zapotřebí další klasifikaci oddělující seskupené články a chyběla by metrika pro posouzení relevance článků mezi sebou. Nabízí se tak vícevrstvá perceptronová síť, síť MADELINE a jiné lineární sítě, u kterých více výstupních neuronů může být aktivních.

Oporu pro určení typu sítě lze najít ve výukovém programu TNeuron dle Teda (2012). Zejména pak úloha nazvaná PERS se zde blíží podobnému cíli. Úloha je z oblasti personalistiky a má na vstupu činnosti, které provádějí pracovníci a kteří tvoří zároveň výstup sítě. Jedná se o síť s učitelem.

Podrobně úlohu popisuje na zjednodušeném příkladu Teda (1997) následovně. V organizaci pracuje 5 pracovníků a je jim přiřazováno 5 činností: opravy motoru, karosérie a elektrického zařízení, podvozku a interiéru. Zpočátku jsou jednotlivé činnosti přiřazovány pracovníkům nahodile, přičemž se provede i hodnocení jejich výkonu. Nula znamená, že požadovaná oprava nebyla úspěšná. Naopak jednička ukazuje na úspěšnou opravu. Po zadání trénovací množiny je síť naučena. Síť umí dávat doporučení a určit tak pracovníka, který dokáže daný úkol splnit nejlépe. V textu je dále popsána možnost rozšíření sítě o hodnocení pracovníků v intervalu 1 až 5 přičemž jednička znamená, že pracovník umí činnost skvěle. Topologie rozšířené sítě je dvojvrstvá. Vstupní vrstvu tvoří činnosti (součet jejich vstupů musí dávat jedničku). Výstupní vrstvu tvoří pracovníci. Naučená síť dokáže i dobře aproximovat. O tom se lze přesvědčit v programu TNeuron, kde lze zadat dotaz i pro hodnoty, které nebyly v trénovací množině. Lze například kombinovat různý poměr na vstupu mezi činnostmi např. 0,3 karosérie a 0,7 elektrické zařízení a síť doporučí vhodného pracovníka na tuto pracovní zakázku.

Řešení úlohy PERS obsahuje návod, jak sestavit neuronovou síť pro řešení zdejší úlohy. Požadavek je, aby na vstupu byli čtenáři a na výstupu články, protože je potřeba, aby čtenáři dostali relevantní články. Vstupní neurony lze prezentovat jako čtenáře, resp. vektor čtenářů,

kdy hodnotící čtenář má jedničku na vstupu a ostatní čtenáři nulu (dimenze vektoru je počet čtenářů). Aktivní výstupní neurony pak ukazují na relevantní články. Situace je zde tedy podobná jako v úloze PERS. Nicméně úvahy v počátcích kapitoly počítají s tím, že neuronová síť bude schopna zpracovat ještě další komponenty, kterými jsou témata a velikost zájmu o témata. Nabízí se přidání skryté vrstvy, která by reprezentovala témata. Velikost zájmu by byla určena váhami v síti. Výhody sítě jsou totožné s výše uvedeným příkladem a jedná se o aproximaci. Tím, že se propojí čtenáři s články a síť se bude učit na základě hodnocení čtenářů, budou se upravovat jednotlivé váhy. Každému novému čtenáři, který samozřejmě bude mít jiné zájmy než ti čtenáři, kteří již články hodnotili, bude doporučen relevantní článek, protože síť dokáže aproximovat jeho požadavky. I při různorodých vstupech je síť schopna doporučit relevantní články. Tím nejenom, že čtenář dostane články odpovídající jeho zájmům, ale také se díky jeho hodnocení zvyšuje synergický efekt neuronové sítě, protože každý další čtenář využívá předchozích hodnocení čtenářů. Použitá neuronová síť bude tedy typu vícevrstvé lineární sítě.

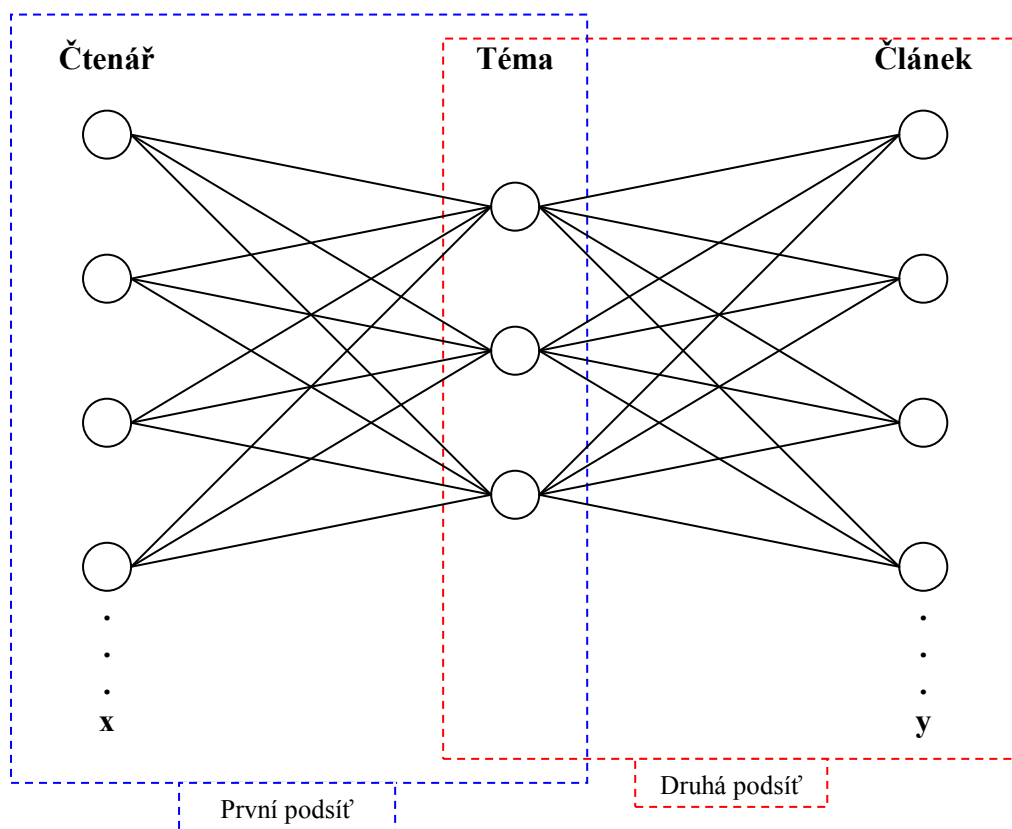
4.2 Model sítě

Jak už bylo naznačeno v předchozí kapitole, zmíněnou úlohu bude řešit jedna síť, v níž budou propojeny všechny komponenty, se kterými se bude pracovat. Zejména pak komponenta čtenář, představující registrovaného uživatele a komponenta článek, která tvoří výstup sítě. Pomocí sítě bude čtenář dostávat relevantní články. Klíčová jsou zde témata, která si budou vybírat samotní čtenáři a zároveň témata budou přidělována automaticky článkům podle jejich obsahu.

I když byla celá předchozí kapitola věnována výběru vhodné neuronové sítě pro řešení stanoveného cíle, výsledný model potřebuje ještě závěrečnou úpravu. Úprava se týká samotných vah mezi neurony, které představují velikosti zájmu o téma a kvalitou článku v rámci tématu. Více o významu vah se uvádí v následujících podkapitolách. Při použití vícevrstvé lineární sítě lze z matematické interpretace modelu vyvodit, že velikost vah se může podstatně lišit v závislosti na původní konfiguraci sítě, přičemž výstupy sítě by byly totožné. Velikost vypočtených vah však může být různá např. hodnota 0,1, ale i 4,5, a obě tyto hodnoty vah mohou aktivovat tentýž výstupní neuron. To je v pořádku, takovým stylem

neuronové sítě pracují a proces adaptace vah je součástí učení sítě. Model však předpokládá, že velikost zájmů čtenáře o téma bude reprezentována právě onou vahou mezi neurony. Aby mohly být váhy upravovány, je pozměněn i adaptační proces, o kterém pojednává samostatná kapitola. V důsledku toho je navržená neuronová síť rozdělena do dvou podsítí, jak je ukázáno na obr. 4.1, které jsou řízeny samostatně. V případě, že jsou čtenáři vybírány relevantní zprávy, jsou tyto dvě podsítě propojeny do jedné, jak je opět vidět na zmiňovaném obrázku. Obě podsítě jsou podobné topologii úlohy PERS. V první podsíti se jedná o propojení čtenářů a témat, v druhé podsíti o propojení témat a článků.

Obr. 4.1 – Model neuronové sítě



Zdroj: vlastní zpracování

Navržený výsledný model tvoří dvě dvouvrstvé lineární sítě, které jsou ve finále propojeny do jedné a tvoří tak jednu vícevrstvou dopřednou lineární síť, která má jako výstupy relevantní články pro konkrétního čtenáře. Použitá síť je typu MADELINE. Aby bylo možné určit relevanci u každého článku, nemá propojená neuronová síť aktivací funkci ve skryté ani výstupní vrstvě.

Potenciál výstupních neuronů celkové sítě se vypočítá tak, že se nejprve vypočte potenciál všech neuronů výstupní vrstvy první podsítě (všech témat) dle vzorce 2.13. Tyto neurony v propojené síti nemají aktivační funkci, takže jejich potenciál se stává vstupem do druhé podsítě. Proces výpočtu je obdobný, pro přehlednost je uvedena jeho podoba:

$$\xi = \sum_{i=1}^n w_i \cdot z_i \quad (4.1)$$

Kde:

- ξ potenciál neuronu výstupní vrstvy,
- n počet vstupních neuronů,
- w_i váha vstupu i ,
- z_i i -tý výstup první podsítě,

Jelikož ani neurony výstupní vrstvy celkové sítě nedisponují aktivační funkcí, tak potenciál se rovná reálnému číslu, které „ohodnocuje“ článek napříč celou sítí podle rovnice 4.1. Při dodržení pravidel pro stanovení vah a učení neuronové sítě, vyšší hodnota potenciálu znamená vyšší relevanci článku.

4.2.1 První podsít'

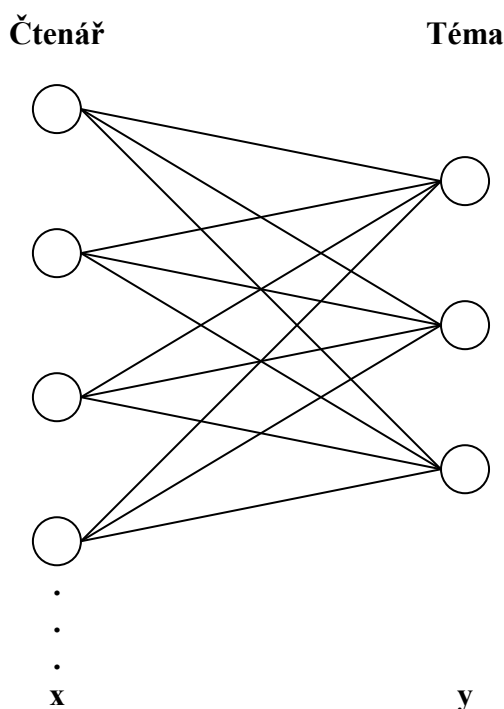
Tato síť zachycuje čtenáře mající své zájmy a témata, která jim odpovídají. Vstupem je čtenář a výstupem je téma. Nastává zde však problém, který je následující. Pokud by si čtenář zvolil všechna témata, pak na výstupu první sítě by byly všechny neurony aktivní. Tyto neurony by po propojení s druhou sítí prakticky simulovaly operátor rovná se. Efektivita sítě by zde ztratila na síle. Ve své podstatě by dávala stejný výsledek, jako kdyby se přidaly atributy zastupující jednotlivá témata do relace Article a poté se uživateli zobrazily (vyfiltrovaly) jen ty záznamy, u nichž atribut splňuje podmínku, tj. nula nebo jedna. Výhoda sítě by se ztratila.

Řešením je využít schopnosti lineární sítě, kterou je aproximace. Aproximace výsledku je ovlivněna volbou aktivní a adaptivní dynamiky. Důležitou roli zde hrají váhy neuronů, které zde mohou představovat velikost zájmu o téma. Díky vhodně zvolené aktivační funkci pak výstup sítě bude úměrný tématu a velikosti zájmu čtenáře o toto téma dle rovnice 2.1. Lze si

představit, že zájem o určité téma je v rozmezí 0 – 100%. Přičemž 100% představuje absolutní zájem o téma.

Na obrázku 4.2 je zobrazena první podsít', která je lineární a dvouvrstvá. Vstupní vrstvu tvoří čtenáři, výstupní vrstvu tvoří témata. Váhy představují velikost zájmu o téma, přičemž platí, že čím vyšší váha, tím vyšší je zájem.

Obr. 4.2 – První podsít'



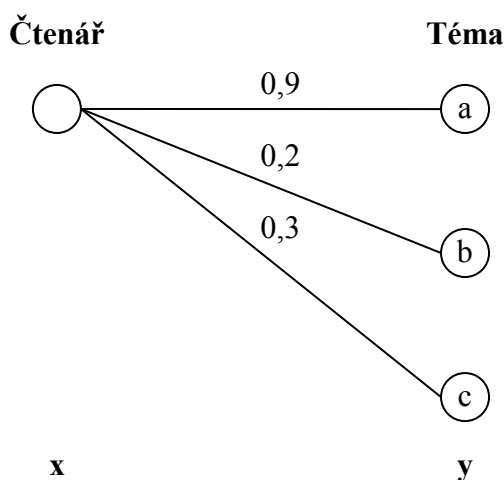
Zdroj: vlastní zpracování

Tato síť by byla nestandardní v tom, že jsou jí přidávány vstupy. Každý nově registrovaný uživatel odpovídá jednomu neuronu ze vstupní vrstvy, tj. čtenáři. Uživatel, který by se sítí právě pracoval, by měl na vstupu jedničku, ostatní čtenáři by měli nulu. Jelikož se ve výsledku podsít' první propojuje s podsítí druhou a vybírají se relevantní články jen pro jednoho konkrétního čtenáře, je lepší, kdyby si každý uživatel držel svoji síť. Odpadne tak povinnost přidat vstupní neuron pro každého nového čtenáře. Model sítě pro každého čtenáře je nazván jako model struktury zájmu.

Model struktury zájmu

Model struktury zájmu se věnuje konkrétnímu čtenáři a je podmnožinou první podsítě. Charakteristika sítě je stejná, liší se pouze organizační dynamika. Model struktury zájmu představuje dvouvrstvou síť, kde na vstupu je pouze jeden neuron, představující konkrétního čtenáře. Hodnota vstupu bude permanentně jedna. Informace obsažené v tomto modelu jsou téma a velikost zájmu o téma. Zájem se bude standardně měnit v procesu učení sítě.

Obr. 4.3 – Model struktury zájmu s váhami



Zdroj: vlastní zpracování

Na obr. 4.3 je vidět jakým způsobem si podsít' udržuje informace o preferencích a zájmech čtenáře. Čtenář má 90% velikost zájmu o téma *a*. V případě tématu *b* je to 20% a u tématu *c* je to 30%. Nejsilnější vazbu má téma *a*, což je pak zohledněno při výběru článku v celkové síti.

4.2.2 Druhá podsít'

Tato podsít' přiřazuje články k tématům. Jedná se o lineární dvouvrstvou síť. První vrstvu tvoří témata, druhou a zároveň výstupní vrstvu tvoří články. Váhy zde představují kvalitu článků v rámci tématu. S příchodem nového článku se přidá do sítě nový výstupní neuron, reprezentující tento článek a provede se ohodnocení článku na témata. Opět platí, že čím vyšší váha, tím více článek odpovídá tématu.

Ohodnocení nově uloženého článku na téma se může zdát jako jednoduchá záležitost. Bohužel tomu tak není. Existují zde dva problémy, které je třeba vyřešit.

Prvním problémem je, že současný IS nedisponuje žádným mechanismem pro rozbor a zpracování textů článků. Dva různé přichozí články nelze tedy od sebe automatizovaně rozeznat na základě jejich obsahu. Jedinou možností je v současnosti ohodnocení článku podle kategorie, ke které patří jeho zdroj. Zdroje jsou do kategorií zařazovány manuálně, takže tato charakteristika může být brána jako spolehlivá. I když mohou existovat případy, kdy články zdroje se jeden od druhého podstatně liší svým zaměřením. Jak klasifikovat jednotlivé články je jiná otázka. Jde samozřejmě vymyslet algoritmus, případně použít jiné metody umělé inteligence pro zpracování textu článku a zejména zjištění témat, kterých se článek týká. Tohle je však úplně jiná otázka a natolik obšírná, že její zpracování by přesáhlo rozměr této diplomové práce. Obecně je model druhé podsítě konstruován tak, aby mohl článek mít více témat, kterých se týká a rovněž u každého tématu sledovat jeho kvalitu vůči článku. Pro zjednodušení řešení je ve zbytku diplomové práce považován vztah 1:1, to znamená, že jeden článek má přiřazeno jen jedno téma. Kvalita k tomuto tématu může být regulována.

Druhý problém se týká samotných vah, určující kvalitu článku vůči tématu. Váhy se budou v čase měnit v procesu adaptace sítě. Protože je uvažována lineární síť s učitelem, budou se váhy adaptovat v závislosti na tom, zda splnily očekávání čtenáře (zda jsou relevantní) anebo nikoliv. Učení bude probíhat na základě ohodnocení uživatele. Adaptace neuronové sítě je více popsána ve zvláštní kapitole. Problém zde tedy nastává u nově přichozích článků, kterým je třeba nastavit počáteční váhy. Pokud by počáteční váhy byly nastaveny konstantou, pak by stačilo jedno negativní ohodnocení čtenářem (které by snížilo váhy) a článek by byl ihned diskvalifikován oproti nově přichozím článkům. Tento postup není ideální, protože ohodnocený článek (i když negativně) má větší vypovídající hodnotu než jakýkoliv nově přichozí článek, který nebyl ještě nikým hodnocen. Druhým řešením je nastavit počáteční konfiguraci neuronové sítě náhodnými čísly, jak bývá zvykem u většiny neuronových sítí. S tímto postupem opět nelze souhlasit. Spolehlivost toho postupu by byla nulová, neboť by nic nezabránilo tomu, aby ve skutečnosti kvalitní článek dostal minimální ohodnocení a naopak. Navíc nelze počítat ani s tím, že by síť konvergovala k optimu pomocí adaptace sítě, protože nelze očekávat, že při současném množství zpráv v IS, by jedna zpráva byla

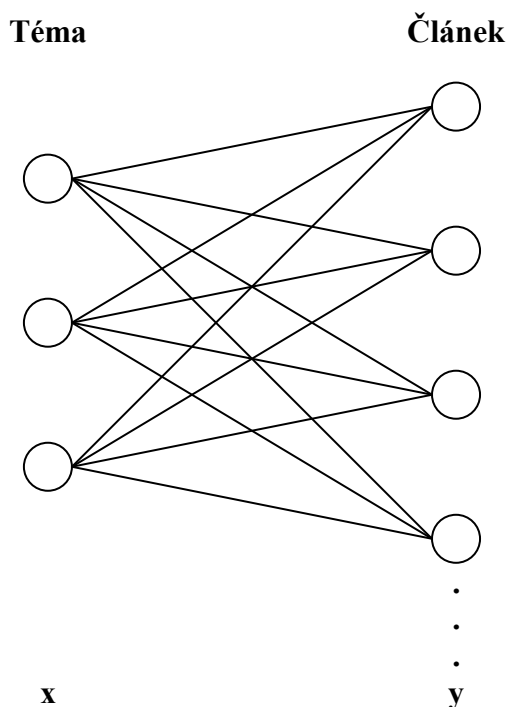
hodnocena vícekrát za sebou, např. 100x. Analyzovaná situace napovídá, že řešení tohoto problému je přijít na to, jak lépe se dá zpřesnit odhad kvality nově přichozího článku.

Jedním z řešení může být další, již třetí podsít', která by propojovala články se zdroji, resp. vydavateli obsahu. Propojením všech tří sítí by se vypočítala kvalita jednotlivých zdrojů. Tato kvalita by se lišila od zdroje ke zdroji a váhy nového článku by se daly nastavit jako kvalita onoho zdroje anebo alespoň v nějakém poměru vůči kvalitě zdroje. Tento způsob je realizovatelný, nicméně ve výsledku je od tohoto způsobu realizace upuštěno.

Podobně jako v předchozím odstavci se tento odstavec zabývá počátečním nastavení vah u nových článků. Způsob uplatnění další podsítě by byl efektivní, kdyby jeden článek pocházel z více zdrojů. To je reálná situace, nicméně současný IS tímto stylem nepracuje. Situaci si lze představit tak, že více zdrojů publikuje stejný článek (ať už se jedná o naprostou shodu textu nebo rozdílné texty nicméně se stejnou obsahovou náplní článků). Nejedná se o výjimečný jev. Některé zdroje opravdu přebírají články z jiných zdrojů, ať už formou pouhého okopírování textu nebo nákupem článků např. od České tiskové kanceláře (přičemž jiné zdroje tak učiní taktéž a shodou okolností oba pak publikují ten samý článek) apod. Pokud by IS disponoval funkcionalitou, která by vyřazovala duplicitní zprávy, pak by třetí podsít' nabyla na významnosti. IS tímto způsobem ale nefunguje. V IS jsou zdroje, kterým náleží jejich články, ať už jsou duplicitní se články jiných zdrojů nebo nikoliv. Nedá se říci, který styl je lepší. Pokud by se eliminovaly duplicity pro uživatele, bylo by to určitě příjemné, je zde pak ale otázka, kterému zdroji přiřadit onen „originál“? Naproti tomu stávající informační systém dovoluje zjistit, které zdroje publikovaly stejný článek, z čeho mohou vyplynout určité statistiky jako například aktuálnost článku (zájem médií o článek), resp. tématu. Jak bylo zmíněno, od třetí podsítě je upuštěno. Pro hodnocení kvality zdroje se použijí současné statistiky zejména analýza čtivosti zdroje a počet článků zdroje.

Protože každý článek představuje jeden výstupní neuron, bude druhá podsít' obrovská. Vazby mezi výstupními neurony nejsou, článek je v rámci této podsítě propojen se všemi tématy. Pro praktickou realizaci je možné články, které jsou v síti, eliminovat tak, aby síť byla menší a tím pádem i rychlejší. Jde však již spíše o technickou záležitost. Odstranit ze sítě se mohou například články, které mají příliš nízké váhy ve všech tématech. Použít se dá rovněž i jednoduché omezení datem zprávy a například tak zpracovávat pomocí sítě články, které jsou jen týden staré.

Obr. 4.4 – Druhá podsít'



Zdroj: vlastní zpracování

4.3 Učení sítě

Učení sítě bude probíhat standardně v procesu adaptace. Protože se jedná o dopřednou síť, adaptace vah bude probíhat jen jedním směrem. Jak již bylo zmíněno v předchozích kapitolách, čtenářem se rozumí registrovaný uživatel. Ten je totiž v IS jednoznačně identifikovaný a bude pro něho v modelu neuronové sítě resp. v modelu struktury zájmu vytvořen jeden neuron, který ho reprezentuje. Tento čtenář má svá oblíbená témata, které chce sledovat a je schopen nastavit velikost zájmu pro každé toto téma. Velikost zájmu reprezentují váhy viz. model struktury zájmu. Dříve než se přistoupí k učení sítě, je třeba vysvětlit, jakým způsobem se budou jednotlivé váhy v podsítích inicializovat.

Pokud jde o první podsít', počáteční váhy si nastaví čtenář ve svém profilu na základě svého uvážení, kde vyjádří velikost zájmu o téma. V průběhu hodnocení se budou automaticky váhy měnit, resp. přizpůsobovat jeho reálným zájmům. Protože třeba během užívání IS se jeho preference změni anebo jeho počáteční nastavení dle jeho uvážení mohlo být chybné.

V případě druhé podsítě je problematika nastavení počátečních vah popsána v kapitole věnující se druhé podsíti. Ve stručnosti lze říci, že pro jednoduchost realizace v praktické části se velikost váhy představující kvalitu článku v rámci tématu stanoví podle kategorie zdroje, ze kterého článek je. To znamená, že článek bude mít pouze jedno téma s proměnlivou kvalitou. Kvalita článku bude nastavena dle současných statistik IS, týkajících se zdroje odrážející jeho zájem u čtenářů.

Samotné učení sítě nastává v okamžiku, kdy čtenář ohodnotí vybraný článek, který mu síť doporučí. Adaptace vah probíhá nezávisle v každé ze dvou podsítí a platí, že se adaptují pouze váhy patřící konkrétnímu čtenáři a článku, který čtenář hodnotí. Hodnocení může nabývat následujících hodnot:

- +1 (článek vyhovuje),
- 0 (čtenář není schopen ohodnotit),
- -1 (článek nevyhovuje).

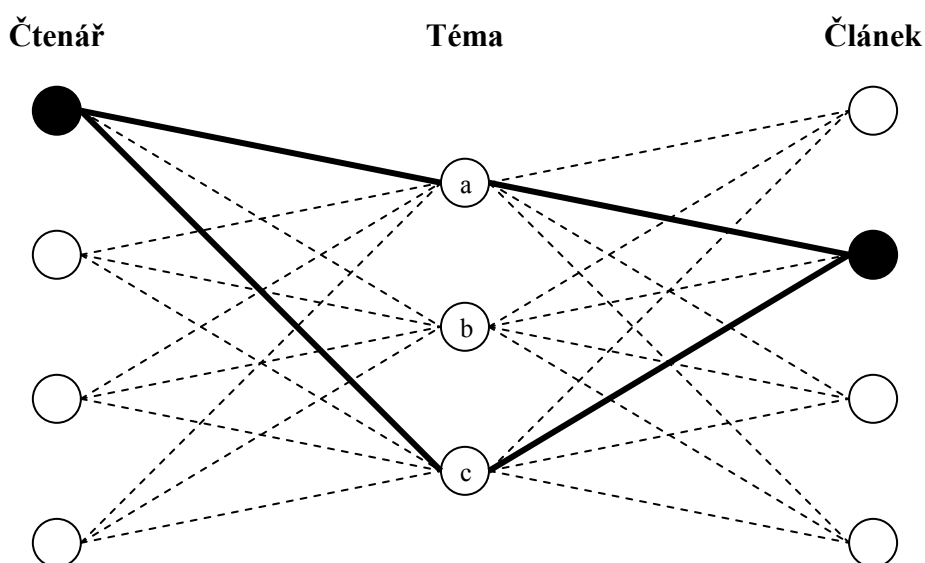
Pokud hodnocení vyhovuje nebo nevyhovuje, provede se adaptace sítě v obou podsítích, přičemž adaptace v každé z nich má svá specifika. Obecně platí, že se vazby posílí, pokud čtenář ohodnotí článek kladně.

Spojovacím prvkem mezi oběma sítěmi jsou témata. Adaptace vah tak probíhá pouze na cestě v síti od čtenáře po článek přes společné téma, jak je vidět na obrázku 4.5. V první podsíti má adaptace vah následující interpretaci. Učením sítě se upravují váhy, které představují velikost zájmu o téma. Když čtenář ohodnotí článek kladně a tento článek má přiřazena určitá témata, pak se preference čtenáře k těmto tématům posílí, v opačném případě sníží. Důvod, proč upravovat váhy v první podsíti je fakt, že preference čtenáře, resp. zájem o témata se může v průběhu času měnit. Kdyby se jeho preference neměnily s každým ohodnocením, pak by síť byla nastavena pouze na základě počátečního nastavení, které čtenář provede na začátku a stala by se statickou. V reálném životě však dochází ke změnám preferencí, zde se tak bude dít automaticky, aniž by čtenář musel úmyslně měnit své nastavení. Uvedený způsob má ještě i vedlejší pozitivní efekt. I když čtenář nastaví na začátku svůj zájem o témata, nemusí být tento zájem pravdivý. Ve skutečnosti ho mohou zajímat jiná témata, o kterých ani neví, že ho zajímají. Zde si je nutno uvědomit, že síť pracuje s množinou témat a nikoliv jen jedním tématem, která čtenáře zajímají a jejichž vzájemné vztahy se ovlivňují. Proto tedy tvrzení, že

čtenář neví, která kombinace zájmů ho může zajímat. A protože se váhy v první podsíti upravují podle společných témat s hodnoceným článkem, dochází zde i k upevnění vah společných témat v obou podsítích.

Lze to vysvětlit tak, že dochází ke kopírování vah z druhé podsítě do první podsítě. Tím jak čtenář hodnotí články, tak se posilují anebo oslabují váhy na cestě od čtenáře ke článku, přes společné téma. Příkladem může být například čtenář, který se zajímá o články s fotbalovou tematikou. V systému budou mimo jiné na výběr témata jako *sport*, *fotbal*, *hřiště*. Intuitivně čtenář nastaví svůj velký zájem pouze o téma *fotbal*. Může existovat situace, kdy většina článků týkajících se fotbalu bude přiřazena ke všem třem zmiňovaným tématům. V tomto případě by čtenář dostal své články, protože zde existuje propojení přes *fotbal*, nicméně pokud by čtenář měl ještě zájem o jiné zcela odlišné téma např. *vaření*, situace by se zkomplikovala. Učení sítě napomáhá tomu, že se v tomto případě vytvoří a posílí vazby i mezi tématy *sport* a *hřiště*. Tím se odhad sítě zpřesní. V úvaze lze jít ještě dále. Může se přihodit, že slovo *fotbal* se stane cenzurovaným ze strany vydavatelů zpravodajského obsahu a bude se místo něho používat slovo *hřiště*, které IS zaindexuje a použije ho tak, že přiřadí toto téma ke článku. Ve výsledku se posílí zájem pouze o téma *hřiště*. Čtenáři budou doporučovány stále relevantní informace ohledně fotbalu, i když ve skutečnosti se jeho model struktury zájmu změnil, což čtenář nemusí ani postřehnout.

Obr. 4.5 – Schéma učení sítě



Zdroj: vlastní zpracování

Na obrázku 4.5 je hodnotící čtenář označen vyplněným kolečkem vlevo a hodnocený článek obdobně vpravo. Předpokladem je, že článek se týká tématu a a c . Pakliže čtenář provede hodnocení, provedou se úpravy vah na vyznačených cestách v síti (plná čára).

U druhé podsítě hodnocení má za důsledek zvýšení či snížení kvality článku v rámci tématu. Pokud je článek ohodnocen kladně, článek je oproti jiným článkům se stejnými tématy zvýhodněn tím, že mu jsou zvýšeny váhy, které pak při výpočtu sítě dle rovnice 4.1 zvyšují jeho rating. Tím, že se zvýšení vah provede i v první podsíti, se vazba na tento článek pro konkrétního uživatele ještě zvýší.

Ve spojitosti s učením sítě je třeba ještě určit velikost koeficientu učení pro rovnici 2.14. Síť MADALINE nevyžaduje speciální úpravy koeficientu jako je tomu u Kohonenovy samoorganizující se mapy, kde by koeficient učení měl s časem klesat. Zde je možné nastavit koeficient učení jako konstantu. Protože význam učení je v každé podsíti různý, bude mít proto každá podsít jiný koeficient učení. Platí, že koeficient učení první podsítě bude menší než koeficient druhé podsítě.

Diplomová práce má za cíl pouze navrhnout model neuronové sítě, její praktická realizace bude však předmětem dalšího rozvoje. V příloze č. 1 je vidět ukázka výstupu z modelu, kde zatím adaptace sítě není softwarově realizována. Jednotlivé články jsou přiřazovány tématům, pro zjednodušení v poměru 1:1 a model struktury zájmu je nastaven podle přiložené tabulky. Každý článek má uveden rating, který tvoří výstup neuronové sítě, jež je generovaný na základě poznatků uvedených výše v této kapitole.

5. Studie proveditelnosti

Studii proveditelnosti popisují například Dennis, Wixom a Roth (2009) jako dokument, ve kterém se popisuje technická, ekonomická a organizační stránka projektu. Smyslem studie proveditelnosti, která se vypracovává před započítáním prací, má být určení, zda je projekt v daných podmínkách realizovatelný či nikoliv. Důležitou částí dokumentu je také popis možných rizik projektu.

„Proveditelnost projektu je ovlivněna řadou faktorů, které je třeba vzít v úvahu. Jde o faktory technického charakteru, které odrážejí schopnost organizace vyvíjet a implementovat daný systém, tedy zkušenost a kvalitu řešitelského týmu, ale i celého implementačního prostředí“, Kaluža (2010, s. 101). Autor dále klade důraz na rizika, legislativu, kterou je třeba při realizaci projektu zohlednit a efektivnost projektu. Lze najít i jiné autory, kteří se zabývají obsahem studie proveditelnosti, přičemž ale jednotlivé části studie se liší. Prioritním cílem diplomové práce není vytvořit studii proveditelnosti, ale navrhnout teoretický model, proto se následující text zabývá jen vybranými částmi této studie, souvisejícími s otázkami využití modelu v praxi.

Již existují programy, které se zabývají neuronovými sítěmi. Jedny jsou komerční, jiné jsou publikovány jako open-source jako například Neuroph, Encog, aj. Jsou zaměřeny na určitý typ úlohy anebo řeší široké spektrum úloh. Dosahují toho tím, že nabízí balíček různých typů neuronových sítí a nechávají na uvážení uživatele, kterou síť pro konkrétní úlohu použít. Některý software dokonce umožňuje kombinovat sítě a nabízí k tomu i vhodné uživatelské rozhraní. Výhodou těchto software je, že nabízí již otestované, stabilní řešení pro různé typy úloh, nabízí možnost rozšíření a disponují většinou i uživatelským rozhraním. Nevýhodou je s tím spojená jejich velikost a špatná integrace do již existujících systémů. V případě použití jen jedné sítě je zapotřebí stáhnout celý program. Další nevýhoda je, že některý software pracuje pouze s daty, které uživatel zadá přes uživatelské rozhraní, neexistuje zde import externích dat např. z databáze. Zdaleka však největší nevýhodou, z pohledu této diplomové práce a již realizovaného informačního systému, je použitá platforma. Většinou se jedná o programovací jazyk Java, popřípadě o jazyk C. Prostředí, ve kterém běží současný informační systém, a kde bude realizován i model neuronové sítě, Javu ani jazyk C nepodporuje. IS využívá webhosting, kde obecně vzato jedinou možností je použití jazyku

PHP. Je to škoda, protože Java je v porovnání s PHP rychlejší, což je důležitá vlastnost v případě, že pracujeme s velkým počtem neuronů či velkým počtem výpočtů.

Situaci lze zlepšit tím, že se například podpora Javy doinstaluje. Bohužel to však nejde na webhostingu. Alternativou mohou být jiné typy služeb jako serverhosting či virtuální hosting. V případě těchto způsobů hostování IS v prostředí internetu je měsíční provoz podstatně dražší než aktuálně využívaný webhosting. Při přechodu na tyto služby by se ztratila mimo jiné i technická podpora současného informačního systému, což by vyžadovalo situaci řešit vlastním personálem případně přikoupením této služby. Změna hostování IS přináší rizika pro stávající IS, což si vyžaduje důkladnější analýzu. Je proto výhodnější zachovat stávající stav a poohlédnout se, jaká existují softwarová řešení založená na jazyce PHP.

Několik takových řešení skutečně existuje. Prvním je projekt ANN (Artificial Neural Network), který podporuje import i export dat. Bohužel pro implementaci zde navrženého modelu neuronové sítě je nevhodný, protože realizuje pouze vícevrstvou perceptronovou síť. Druhým testovaným produktem je NeuralMesh. Řeší pouze binární lineární síť s aktivační funkcí sigmoida. Výhodou je, že disponuje přehledným uživatelským rozhraním a plně spolupracuje s databází, kde ukládá informace o konfiguraci sítě. Ani tento produkt však není vhodný pro implementaci zdejšího modelu neuronové sítě. Dalším je projekt FANN (Fast Artificial Neural Network). Jedná se o propracovaný program, který i když je napsaný v jazyce C, tak nabízí až 15 rozšíření včetně PHP, které používají jeho jednotnou API. Tento program nabízí vícevrstvou lineární síť s algoritmem backpropagation, kaskádovou síť a více druhů aktivačních funkcí. Posledně zmiňovaný program se ze všech zmiňovaných požadovanému řešení přibližuje asi nejvíce. Nicméně i když je zde možnost měnit různá nastavení, bariéra v podobě podpory jazyku C zůstává. S ohledem na výše zmíněné informace a charakter navrženého modelu neuronové sítě, nebude model realizován s využitím software třetích stran, ale bude naprogramován.

Tím, že se síť skládá z dvou podsítí, je těžké hledat software, který tuto a ještě další specifika zvládne vyřešit. Jak bylo zmíněno v kapitole 4.1, jednotlivé neurony reprezentují čtenáře, témata a články. Servery s webhostingem obvykle mají omezení kapacity velikosti operační paměti, rychlosti procesoru, různá velikost databáze a úložného prostoru, aj. Poskytovatel webhostingu, kde bude neuronová síť implementována, tato omezení nemá, zato však existují

jiná, které je třeba brát v úvahu. Jedná se zejména o maximální velikost paměti pro jeden skript a maximální dobu vykonávání skriptu. První parametr je nastaven na hodnotu 30 MB a druhý parametr na 5 minut. Pokud bychom uvažovali, že reprezentace jednoho neuronu by spotřebovala 100 Byte, pak by skript mohl obsloužit cca 300 tisíc neuronů. Jedná se o odhad, protože zatím přesně nelze říct, kolik bytů paměti ve skutečnosti zabere jeden neuron. Navíc do celkového limitu se započítává i paměť, která je spotřebovaná na režii celého skriptu, tedy například i paměť spotřebovanou při připojení k databázi. Aktuálně se v IS nenachází více než 100 tisíc článků, takže lze říci, že limit pro paměť je dostatečný. Protože jsou do systému přidávány nové zdroje, je měněna jejich doba archivace apod., nelze zaručit, že paměť bude dostatečná i v budoucnu. Nicméně je zde určitá rezerva, takže neuronovou síť lze realizovat tak, že se celá načte do operační paměti a provedou se v ní výpočty. Tím je zaručena vysoká rychlost zpracování. Bez předchozího testování však nelze odhadnout, jestli se toto zpracování stihne do 5 minut, což je druhá podmínka. S ohledem na budoucí rozvoj IS i neuronové sítě, je dle mého uvažování vhodnější použít jiný způsob implementace. Může jít o paralelní zpracování v rámci jedné vrstvy, jak uvádí Volná (2009) anebo způsob distribuovaného zpracování popsany v dalších odstavcích.

Protože neuronová síť bude potřebovat velký výpočetní výkon, je vhodné si ujasnit, kdy k tomu bude docházet. Práci neuronové sítě lze předpokládat v těchto činnostech:

1. nastavení modelu struktury zájmu čtenáře,
2. ohodnocení nově uloženého článku,
3. ohodnocení článku uživatelem,
4. doporučení relevantních článků uživateli.

V kapitole 4 je vysvětleno, proč je model rozdělen na dvě podsítě a jak tyto podsítě spolupracují. Toto rozdělení sítě může být využito i při návrhu implementace. Každá podsít' může být realizována jako samostatná síť bez vazby na celou síť. Výpočty by se prováděly jen v rámci této podsítě, což by mělo za následek menší paměťové nároky a nároky na výpočetní výkon, což by zlepšilo situaci popsanou výše. Jde tak uvažovat o činnostech 1 a 2. V případě činností 3 a 4 je zapotřebí interakce obou podsítí, která může být realizována například voláním metod mezi objekty. Přičemž nárok na výkon zůstává pořád rozdělen mezi obě sítě.

V případě hardwarových nároků je model neuronové sítě opět omezen již existujícím prostředím. Obecně pro algoritmy umělé inteligence je zapotřebí dostatečné množství

operační paměti. V té se provádí velké množství výpočtů. Další důležitou komponentou je rychlost procesoru, která by měla být opět co nejvyšší. Naproti tomu grafická karta není vůbec zapotřebí, neboť se výpočty neuronové sítě neprezentují v tomto případě vizuálně. A pokud máme vstupní data na disku, není potřeba ani síťového rozhraní. Disk, ze kterého se čtou data, by měl mít rychlou přístupovou dobu. V případě poskytovatelů webhostingových a podobných služeb je zapotřebí, aby spojení mezi databázovým a aplikačním serverem bylo dostatečně rychlé v případě, že se tyto nacházejí na různých fyzických zařízeních.

Přínosy neuronové sítě se očekávají zejména ve zkvalitnění služeb, resp. ztraktivnění obsahu, což sníží fluktuaci současných uživatelů a eventuálně zvýší počet nově registrovaných uživatelů na základě dobré reference. To zlepší situaci na trhu a může přinést i konkurenční výhodu. Vedlejší efekt je očekáván v přesnějším určení zájmů uživatelů, který může správce webu využít tak, že bude rozšiřovat nabídku zdrojů nabízející žádanou tematiku, odstraňovat nepotřebné zdroje či snižovat jejich dobu archivace, což v celkovém důsledku povede k větší efektivitě současného IS. U zdrojů, které budou hodně využívány, je možné uvažovat o užší spolupráci s vydavatelem obsahu. Výhoda je zřejmá i při registraci nového uživatele. Doposud si sám musel určit, která kategorie, zdroje a podobně ho zajímají. Investoval určitý svůj čas do personalizace stránek. V případě, že by při registraci vyplnil dotazník (který by byl sestaven na základě poznatků o uživateli z neuronové sítě), byl by mu ihned přiřazen pravděpodobný model struktury zájmu a byly by mu doporučeny relevantní články, resp. zdroje. Uživatel by nemusel zkraje nic nastavovat, pouze by upravil to, co mu bylo doporučeno. Vše by se zrychlilo, což uživatele více uspokojí.

5.1 Způsob prezentace neuronové sítě uživateli

Úspěšnost zavedení nové funkcionality v podobě neuronové sítě závisí na mnoha faktorech. Pominou-li se klasická rizika projektu, marketing, kvalita dat, apod. je dle mého názoru klíčové uživatelské rozhraní. Je třeba si uvědomit, že se jedná o novou věc. Čeští čtenáři, resp. uživatelé, buď funkcionality vůbec neznají, případně na ni nejsou dostatečně zvyklí. Neví, co od funkcionality očekávat. Tohle je velký problém, protože tímto můžeme uživatele odradit a to natolik, že nebude IS využívat vůbec. Anebo funkcionality nebude používána z důvodu strachu uživatele, jeho neznalosti, apod. Nabízí se dvě možnosti řešení. První je, že se

funkcionalita prezentuje uživateli s velkou noblesou a se vším, co k tomu patří. V takovém případě se vytvoří dostatečně názorný manuál, zavedení se podpoří reklamou, stávajícím uživatelům se zašlou informační e-maily, apod. Druhý způsob zavedení je opačný. Uživatel se o neuronové síti respektive její fyzické existenci vůbec nedozví. Vše, až na hodnocení článků, bude skryto. Zmiňované hodnocení článků uživatelé znají z jiných serverů a jsou na ně již zvyklí, takže toto činit problémy nebude. Nová funkcionalita bude skrytě zakomponována do stávajícího systému, přičemž ji uživatelé budou využívat, aniž by tak záměrně činili. Ve výsledku by pak vše vypadalo tak, že stávající výstupní sestavy by měly ještě jeden parametr, kterým je vypočtená relevance článku, přičemž uživatel by o tomto parametru nevěděl. Který způsob upřednostnit, je otázkou.

Další otázkou je, jak nazvat novou funkcionalitu. Může se jednat např. o *relevantní články*, *umělá inteligence*, *personalizace*, *neuronová síť*. Ať už označení je více či méně přitažlivé, asi nejlepší způsob je, podívat se ke konkurenci, která již s umělou inteligencí zkušenosti má. Mimo firmy fungující v profesionálním monitoringu tisku, které umělou inteligenci používají ve velké míře jako například Newton Media, a. s., pak na českém trhu spustila v roce 2011 mediální agentura OMD, a. s., provoz obdobného serveru na adrese chytrenoviny.cz. Tento informační systém také nabízí relevantní články. Při registraci si vyberete svou sociální skupinu např. student a jsou vám předvybrány kategorie a zdroje, které by vás mohly zajímat. Autoři rovněž popisují propojení s facebook.com v podobě doporučování vhodných článků svých přátel. Zda je systém skutečně postaven na metodách umělé inteligence, však netuším. Nicméně zde byla provedena velká reklamní kampaň vyzdvihující onu odlišnost od ostatních serverů s důrazem na slova *chytrý*, *intelligentní*. Na tomto serveru je tedy funkcionalita řádně prezentována, dokonce je na ní založena hlavní konkurenční výhoda. Uživatel při první návštěvě stránek ihned vytuší, že je zde jiné ovládání, které se řídí jinými principy.

Jiným příkladem může být news.google.com společnosti Gogole, Inc., která rovněž obdobnou funkcionalitu zpřístupnila, ovšem jako doplněk běžně vyhlízejících webových stránek ve formuláři pod názvem *personalizace*. Zde lze pomocí posuvníků nastavit velikost zájmu o kategorie a jde i přidat vlastní zájmy, resp. klíčová slova. U formuláře je pouze nenápadný odkaz na nápovědu.

Výše uvedené příklady ukazují na různorodost prezentace a začlenění nové funkcionality. Ať už se zvolí první či druhý způsob, v žádném případě by to uživatele nemělo odradit. Kvalitní

nápověda je na místě a člověk spravující technickou podporu by měl být proškolen, aby uměl vysvětlit přínosy neuronové sítě uživateli. Je třeba mít na paměti, že kvalita neuronové sítě závisí na velkém množství zpětných vazeb, které budou dávat uživatelé tím, že články budou hodnotit. Pokud se nepodaří zapojit uživatele do tohoto procesu, pak výhody, které poskytuje toto řešení, se omezí.

5.2 Rizika projektu

Pro metody umělé inteligence je charakteristické, že pracují s velkým objemem dat. Ne jinak je tomu i v případě navrženého modelu neuronové sítě, který bude pracovat s daty z informačního systému, ve kterém je v současné době několik desítek tisíc článků. Jsou zde kladeny velké požadavky na výpočetní výkon. Během testování se může zjistit, že hardwarové či softwarové prostředí zmiňované v předchozí kapitole není dostatečné. Problémem také je, že například doporučení relevantních článků je možné až po zpracování celé neuronové sítě. V případě, že proces zpracování bude násilně přerušen (což se může reálně stát například tehdy, když se překročí maximální doba vykonávání skriptu), nebude část provedených výpočtů použitelná. Zabezpečit se proti tomuto riziku lze dostatečným testováním ve vývojové fázi a následným výběrem vhodného způsobu implementace. Logování chyb by taktéž mohlo přispět k lepšímu zvládnutí řízení rizik. Eventuálně lze eliminovat počet zahrnutých článků, například zadáním rozmezí v datu, případně zpracovat jen články, které uživatel ještě nečetl či nehodnotil, apod.

Uživatelé musí systém aktivně využívat, aby se dostavily jeho výsledky. Primární funkcí lineární sítě je aproximace, která je tím přesnější, čím častěji se síť adaptuje. Adaptaci provádí uživatelé prostřednictvím hodnocení článků. Hodnocení jednoho uživatele může ovlivnit výsledky, které jsou předány jiným uživatelům se stejnými nebo podobnými váhami. Dochází tak k synergickému efektu neuronové sítě. Je důležité zapojit co nejvíce uživatelů do procesu hodnocení. Může se také stát, že uživatelé sice funkcionalitu budou používat, nicméně nebudou hodnotit, ať už z nějakých vnitřních přesvědčení anebo proto, že je to bude stát úsilí navíc. Aby se předešlo této situaci, je dobré zamyslet se nad jinými způsoby zjištění hodnocení uživatele. Příkladem může být uživatel, který doporučí článek svému známému pomocí tlačítka. Dá se předpokládat, že jeho hodnocení tohoto článku by bylo pozitivní.

Zejména při prvním zavedení neuronové sítě bude problém udržet zájem uživatelů využívat funkcionalitu, neboť výsledky nebudou dostatečně relevantní. Je to dáno tím, že váhy mezi neurony budou pouze inicializovány dle kapitoly 4.2 a nikoliv adaptovány a tak hrozí, že nebudou nastaveny optimálně. Zejména váhy druhé podsítě reprezentující kvalitu článku v rámci tématu se mohou velmi lišit. Stejně jako u první podsítě se musí zajistit, aby váhy neklesly na nulu, protože by byly takto ohodnocené články automaticky vyřazeny. I s vývojem sítě v čase zde mohou nastat problémy. Například se může stát, že určitý zdroj bude neustále preferován sítí, i když ho uživatelé budou hodnotit negativně. Příčinou mohou být koeficienty učení nebo statistika, ze které jsou počáteční váhy vypočítávány. Je nutno pečlivě definovat hodnoty koeficientů učení. Příliš vysoké číslo může vést až k přeučení neuronové sítě, které je charakteristické kopírováním vstupů. Naproti tomu malé číslo nemusí být dostatečné natolik, aby mělo patřičný vliv. Kvalita vstupních dat také ovlivňuje neuronovou síť. Pro zjednodušení má druhá podsít' pouze jeden vstupní neuron, odpovídající kategorii, do které je přiřazen zdroj článku. Může se však stát, že zdroj nebude respektovat onu kategorii a články budou jiného charakteru. Následně dojde k záměně a hodnocení nebude adekvátní.

6. Závěr

Před započítím prací jsem prostudoval dostupnou literaturu, která je uvedena v seznamu v závěru diplomové práce. Na základě těchto poznatků jsem sepsal kapitolu popisující teorii neuronových sítí. Pro lepší orientaci v problematice jsem nejprve uvedl všeobecné charakteristiky neuronových sítí, zejména pak jednoduchý model neuronu a jeho vazbu na biologický neuron. Protože jsou jednotlivé typy neuronových sítí odlišné z pohledu organizační, aktivní a adaptivní dynamiky, následovala část vysvětlující i tyto pojmy. Diplomová práce se zabývá návrhem modelu neuronové sítě, proto jsem do teoretické části zahrnul i konkrétní typy vybraných neuronových sítí, u nichž jsem se domníval, že mohou být použity. Jednalo se o Perceptronovou síť, Kohonenovu samoorganizující se mapu a síť MADALINE. Jsou to jedny z nejpoužívanějších sítí. Popsal jsem proto jejich možný způsob užití a všechny tři dynamiky.

Jelikož cíl diplomové práce je jen obecný, bylo třeba ho blíže specifikovat. Naskytla se možnost použít data z informačního systému, který se zabývá sběrem zpravodajských článků. Popsal jsem problematiku zpravodajství obecně i konkrétní informační systém, který bude poskytovat vstupní data neuronové sítě. Na základě tohoto popisu jsem navrhl konkrétní úlohy a následně vybral jednu, kterou bude neuronová síť řešit. Zhodnotil jsem možnost použití jednotlivých typů neuronových sítí pro řešení dané problematiky a navrhnul jsem model, který řeší vytýčený cíl. Následoval detailní popis neuronové sítě a její organizační dynamiku jsem zakreslil. Z analýzy problému vyplynulo, že ve skutečnosti se síť bude dělit na dvě podsítě. První podsít' zachycuje vztah čtenáře a témat, druhá podsít' pak vztah mezi tématy a články. Každá podsít' je detailně popsána. Ve výsledku budou obě sítě spolupracovat a spojí se do jedné sítě. Samostatnou podkapitolu jsem věnoval adaptaci sítě, resp. podsítí. Průběh adaptace je zobrazen rovněž graficky a je uveden význam jednotlivých spojení mezi neurony.

U navrženého modelu jsem zhodnotil možnosti praktického využití a rizika jeho nasazení. Zaměřil jsem se na hardwarové a softwarové nároky. Uvedl jsem příklady již existujících software, které se zabývají neuronovými sítěmi, a okomentoval jejich použitelnost. Blíže jsem popsal prostředí, ve kterém běží existující informační systém a kde bude model neuronové sítě realizován. Nastínil jsem způsoby implementace a zhodnotil přínosy modelu. V závěru

jsem se věnoval způsobu prezentace neuronové sítě coby prvku umělé inteligence pro uživatele.

Cílem diplomové práce bylo prozkoumat možnosti neuronových sítí a navrhnout model neuronové sítě, který by dokázal uživateli doporučit relevantní informaci podle jeho profilu.

Vytyčený cíl se podařilo splnit, model neuronové sítě byl sestaven v kapitole čtvrté.

Pokud se model neuronové sítě podaří uvést do praxe, přinese to uživateli informačního systému zkvalitnění služeb, resp. zatraktivnění obsahu. Články generované informačním systémem dostanou nový rozměr, kterým je relevance. Relevance bude pro každého uživatele jiná a to taková, jakou si nastaví. Neuronová síť je v tomto případě skvělým pomocníkem. Dostatečné množství hodnotících uživatelů totiž vytvoří synergický efekt, ze kterého může čerpat každý další uživatel, který tak zajistí ocenění relevantní informace.

Seznam použité literatury

Knižní zdroje:

- (1) BERKA, Petr. *Intelligentní systémy*. Praha: Oeconomica, 2008. 100 s. ISBN 978-80-245-1436-9.
- (2) CONOLLY Thomas, Carolyn BEGG a Richard HOLOWCZAK. *Mistrovství databáze, profesionální průvodce tvorbou efektivních databází*. Brno: Computer Press, 2009. 550 s. ISBN 978-80-251-2328-7.
- (3) DENNIS Alan, Barbara WIXOM a Roberta ROTH. *Systems analysis & design*. Hoboken: Wiley, 2009, p. 530. ISBN 978-0-470-22854-8.
- (4) KALUŽA Jindřich. *Informační systémy pro strategické řízení*. Ostrava: VŠB-Technická univerzita Ostrava, 2010. 145 s. ISBN 978-80-248-2280-8.
- (5) MARČEK, Dušan a Milan MARČEK. *Neurónové siete a ich aplikácie*. Žilina: Edis, 2006. 214 s. ISBN 80-8070-497-X.
- (6) OLEJ, Vladimír a Petr HÁJEK. *Úvod do umělé inteligence*. Pardubice: Univerzita Pardubice, 2010. 98 s. ISBN 978-80-395-307-2.
- (7) RUSSELL, Start. *Artificial Intelligence: A Modern Approach*. 3rd ed. Prentice Hall, 2009. p. 1152. ISBN 978-0136042594.
- (8) VONDRÁK, Ivo. *Umělá inteligence a neuronové sítě*. 3. vyd. Ostrava: VŠB-TUO, 2009. 140 s. ISBN 978-80-248-1981-5.
- (9) ŠÍMA, Jiří a Roman NERUDA. *Teoretické otázky neuronových sítí*. 1. vyd. Praha: MATFYZPRESS 1996. 196 s. ISBN 80-85863-18-9.
- (10) VOLNÁ, Eva. *Neuronové sítě 1*. Ostrava: Ostravská univerzita, 2008a. 86 s.
- (11) VOLNÁ, Eva. *Neuronové sítě 2*. Ostrava: Ostravská univerzita, 2008b. 83 s.
- (12) VRÁNA, Jakub. *1001 tipů a triků pro PHP*. Brno: Computer Press, 2010. 442 s. ISBN 978-80-251-2940-1.
- (13) ZULÁK, Ondřej. *On-line agregátor zpráv*. Ostrava, 2008. Bakalářská práce. Vysoká škola báňská – Technická univerzita Ostrava, Fakulta elektrotechniky a informatiky, Katedra informatiky.

Odborné časopisy, sborníky:

- (14) ŠMEJKAL, Pavel. Zvládnou v budoucnu počítače přemýšlet? Zatím to umějí jen trochu! *21. století*. 2012, č. 2, s. 58-59. ISSN 1214-1097.

- (15) TEDA, Jaroslav a Jiří SCHINDLER. *Personálního zabezpečení servisních činností metodou neuronových sítí. MOSISS'97. Díl 2.* Ostrava: VŠB-TUO, 1997. 200 s. ISBN 80-85988-16-X.
- (16) VOLNÁ, Eva. Vícevrstvá neuronová síť jako univerzální aproximátor. *Automatizace*. 2009, č. 11, s. 658-660. ISSN 0005-125X.

Internetové zdroje:

- (17) SKOPAL, Tomáš. *Neuronové sítě a Information Retrieval. Amphora Research Group* [online]. 2011. Dostupné z: <http://www.cs.vsb.cz/arg/techreports/neural.pdf>.
- (18) Slovník cizích slov.cz. [online]. 2012. Dostupný z: http://www.slovník-cizich-slov.cz/relevantni_informace.html.
- (19) TEDA, Jaroslav. *Neuronové sítě*. [online]. 2012. Dostupné z: <http://www.volny.cz/jaroslav.teda/>.

Seznam zkratek

ADALINE	Adaptive Linear Element
ANN	Artificial Neural Network
API	Application Programming Interface
Atom	Atom Syndication Format
aj.	a jiné
apod.	a podobně
DFD	Data Flow Diagram (Diagram datových toků)
FANN	Fast Artificial Neural Network
INNS	International Neural Network Society
IS	Informační systém
Java	Programovací jazyk
kyborg	Kybernetický organismus
MADALINE	Multiple Adaptive Linear Element
MB	mega byte
např.	například
NUTS	Nomenklatura územních statistických jednotek
resp.	respektive
PHP	Skriptovací programovací jazyk
RSS	Really Simple Syndication
USA	United States of America
UTF-8	8 bitový způsob kódování znaků
XML	Extensible Markup Language

Prohlášení o využití výsledků diplomové práce

Prohlašuji, že

- jsem byl seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. – autorský zákon, zejména § 35 – užití díla v rámci občanských a náboženských obřadů, v rámci školních představení a užití díla školního a § 60 – školní dílo;
- beru na vědomí, že Vysoká škola báňská – Technická univerzita Ostrava (dále jen VŠB-TUO) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užít (§ 35 odst. 3);
- souhlasím s tím, že diplomová práce bude v elektronické podobě archivována v Ústřední knihovně VŠB-TUO a jeden výtisk bude uložen u vedoucího diplomové práce. Souhlasím s tím, že bibliografické údaje o diplomové práci budou zveřejněny v informačním systému VŠB-TUO;
- bylo sjednáno, že s VŠB-TUO, v případě zájmu z její strany, uzavřu licenční smlouvu s oprávněním užít dílo v rozsahu § 12 odst. 4 autorského zákona;
- bylo sjednáno, že užít své dílo, diplomovou práci, nebo poskytnout licenci k jejímu využití mohu jen se souhlasem VŠB-TUO, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly VŠB-TUO na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Ostravě dne

.....
jméno a příjmení studenta

Seznam příloh

Příloha č. 1 Příklad výstupu neuronové sítě

Příklad výstupu neuronové sítě

Téma	Velikost zájmu
Ekonomika	100 %
Sport	80 %
Domácí zpravodajství	20 %
Ostatní	méně než 20 %

Zdroj: část modelu struktury zájmu čtenáře (test)

rating: 103 | Daně není nutné zvyšovat. Stačí zajistit ... (EuroZprávy.cz - Ekonomika, finance) 24.4.2012 09:46

rating: 102 | Facebook už má 901 milionů uživatelů, nabírá... (iHNed.cz - Investice) 24.4.2012 09:26

rating: 102 | Facebooku klesl kvůli agresivní expanzi zisk (Týden.cz - Ekonomika) 24.4.2012 09:06

rating: 102 | Předáci považují demonstraci za vítězství.... (E15.cz) 24.4.2012 10:16

rating: 102 | Rozhovor týdne: Steve Jobs byl jedinečný. Vy... (Ekonom) 24.4.2012 10:06

rating: 102 | ČSÚ: Důvěra v českou ekonomiku se v dubnu snížila (E15.cz) 24.4.2012 09:46

rating: 102 | ŘSD utrácí miliardy protizákonně, uvádí kontrola (E15.cz) 24.4.2012 09:46

rating: 102 | Výdaje Googlu na lobbying rostou, v prvním... (iHNed.cz - Investice) 24.4.2012 09:26

rating: 102 | Česko vede v dostupnosti údajů o firmách (E15.cz) 24.4.2012 08:46

rating: 102 | Fico v Praze: Při obhajobě jádra spojíme síly (E15.cz) 24.4.2012 08:46

rating: 102 | Podpora Nabucca se drolí: projekt hodlá... (E15.cz) 24.4.2012 10:16

rating: 102 | Policie vyšetřuje kvůli machinacím poslance VV... (E15.cz) 24.4.2012 10:16

rating: 83 | Češi na Euru vypadnou s Holanďany, spočítali... (Aktuálně.cz - sport) 24.4.2012 08:56

rating: 82 | Rooneyho otec byl zbaven podezření z... (Sport.cz) 24.4.2012 10:06

rating: 82 | Veselý ani jednou nemnul a 16 body si... (Lidové noviny - Sport Report) 24.4.2012 09:46

rating: 82 | Slovan Bratislava by měl v KHL vést český kouč... (Lidové noviny - Sport Report) 24.4.2012 10:06

rating: 82 | Populární vousáč v turbanu ukončil maratonskou... (iDNES.cz Sport) 24.4.2012 09:16

rating: 82 | Štěpánek čeká v Brně zkoušku ohněm. A... (Sport.cz) 24.4.2012 09:46

rating: 82 | Jarošíkova kalvárie. Na podzim Spartu za... (Sport.cz) 24.4.2012 08:56

rating: 82 | Legendární hokejový gólman Tretjak slaví šedesátku (Sport.cz) 24.4.2012 09:46

rating: 82 | 1 výhra = extraliga. Chomutov může roztočit... (iDNES.cz Sport) 24.4.2012 09:46

rating: 32 | Řidič naházal značky se zákazem na hromadu,... (iDNES.cz Zprávy) 24.4.2012 09:06

rating: 32 | Důchodkyně hnala podvodníky holí, venku vpadli... (iDNES.cz Zprávy) 24.4.2012 09:36

rating: 27 | Lesy ČR vykácely na Blanensku 130 let starý... (Mediafax) 24.4.2012 09:36